

In defence of animal research

The use of animals in laboratory research needs championing more than ever. Those defending it need to reflect discussions within the research community and engage positively with issues of public concern.

Animal research is essential for scientific and medical progress, and animal sacrifice is regrettably sometimes needed in the pursuit of human needs. But the conduct of such research needs continual reassessment as scientific knowledge grows and ethical and societal perspectives evolve. As animal-rights activists become more sophisticated in their attempts to hobble and ultimately end the use of laboratory animals, it is vital to preserve and strengthen public confidence.

At issue in the United States are two initiatives intended to toughen protection for research animals. One would, for the first time, define 'distress' under the 1966 Animal Welfare Act. It would also institute a new reporting system aimed at better quantifying the intensity and duration of pain and distress. (The current reporting system contains no scale for measuring these.) In July the US Department of Agriculture (USDA), which enforces the act, asked for comments on how to go about making these changes. The deadline is 7 November.

The second initiative would add rats, mice and birds to the protection of the Animal Welfare Act. USDA has used this law's ambiguous wording to avoid monitoring rodents and birds, arguing that the government does not have the funds for this. Rats and mice comprise some 95% of lab animals, and this will doubtless increase as the boom in transgenic mice continues. Animal-rights activists last year brought a suit against the department for "arbitrarily" excluding the animals. Perhaps anticipating the weakness of its legal case, USDA last week agreed to settle the suit by initiating the administrative process that would probably result in extending the law.

Both moves to expand the law's reach have elicited protest from research advocates in Washington, led by the Association of American Medical Colleges (AAMC), the National Association for Biomedical Research (NABR) and the Federation of American Societies for Experimental Biology (FASEB). These groups argue that such changes would put an onerous, expensive new burden on researchers and their institutions without improving animal protection one whit.

On the issue of rodents and birds, their arguments found an important ally in the US Senate last week. As the lawsuit was being set-

tled, AAMC asked leaders from the University of Mississippi Medical Center to lobby Thad Cochran, a Republican Mississippi senator, who chairs the Senate agriculture spending subcommittee. He responded by altering the bill financing USDA in 2001, forbidding it to settle the lawsuit with the activists. A joint House-Senate committee unanimously approved the amendment without debate. At the least, the research lobby has succeeded in winning a one-year reprieve.

The impression given by groups such as AAMC, NABR and FASEB is that researchers are united in opposing the changes. But the respected American Association for Laboratory Animal Science supports the coverage of rats and mice, calling their exclusion from the law's protection "ethically indefensible". Consider, too, a survey published last year in *Lab Animal* (Vol. 28, issue 6, pp. 38–40). It canvassed 491 members of animal care and use committees, who oversee the use of animals under the Animal Welfare Act, and nearly two-thirds of whom are themselves animal researchers. Fully 73% felt the act should cover rodents. As the study authors concluded, it would be a distortion to present the debate as a conflict between animal researchers and animal protectionists.

Some of the research lobby's arguments verge on the reactionary. As FASEB points out, a definition of distress being considering by USDA is vague and could lead to widely varying interpretations. But the answer is surely not to dispense with a definition, as FASEB urges. AAMC, meanwhile, points to the "millions of dollars and uncountable hours" that would have been wasted had USDA's "misguided" attempt to settle the rodent lawsuit succeeded. It argues that scientists already apply the best possible care to rodents because good science and Public Health Service regulations require it (see also page 671). But the act would have the force of law. It would add unannounced inspections. Big universities may have nothing to hide, but what about the many smaller institutions and private biotechnology companies that use only rodents and so haven't yet fallen within the reach of the law?

Costs and practical burdens are indeed issues to be confronted. But research lobbyists who have often stated that it is a privilege to use lab animals now risk giving the impression that some of them consider it a right. If that continues, research could suffer. ■

Blurred budget is bad news

Spain's government should stop shooting itself in the foot.

The Spanish government's announcement of an 11.3% increase in its research and development budget should have been good news. Coming on top of the launch of a Ministry of Science and Technology, statements of priority for R&D by José Maria Aznar, the prime minister, and a four-year plan to increase R&D spending from 0.9 to 2% of gross national product, 2000 looked like an *annus mirabilis* for Spanish science.

But scientists were suspicious. Complaints about the true distribution of R&D budgets have been common since 1998, when it became clear that a substantial amount of the annual spend was set aside for military research. But distrust has grown as the government has resisted publishing the military and civilian allocations.

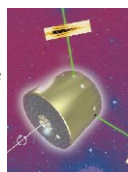
Belatedly, next year's position is now all too clear (see page 664). The military research spend will increase disproportionately. And to add insult to injury, secrecy rules again. Only reluctantly was science minister Ramón Marimón forced to admit that more than 50% of the 2001 budget will be devoted to military R&D.

In seeking to hide the military figures, the government is doing neither its citizens nor itself any favours. Why exacerbate distrust in this way? The United States, France and the United Kingdom make a clear distinction between civilian and military spending, and that information is easily accessible to all. To restore disenchanted scientists' trust in the government, Spain's government needs to boost civil science, significantly and transparently. ■





Rapid rodent
Public-private mouse
genome project steps
up a gear
p663



Lift-off
Germany chooses
space mission to
map the heavens
p665



Bowled over
Ig Nobels take a
sideways look at
scientific endeavour
p665



Mummy?
Rare ox cloned and
implanted in cow
surrogate mother
p666

Medicine Nobel goes to raiders of the brain's chemical secrets

Alison Abbott

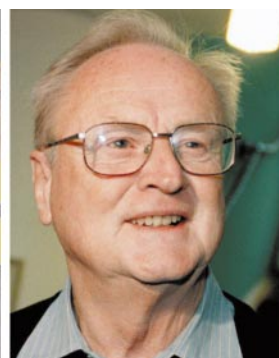
In 1990, the film *Awakenings* won Robert De Niro an Oscar nomination for his portrayal of a patient with a severe form of Parkinson's disease, who was released from his trance-like state by the drug L-DOPA. This week, the pharmacologist who made the treatment possible is one of three researchers to win the Nobel Prize in Physiology or Medicine for their pioneering discoveries about how signals are transmitted between nerve cells.

Arvid Carlsson, of the University of Gothenburg in Sweden, is honoured for showing that dopamine is a neurotransmitter in the brain, and that lack of this chemical causes the symptoms, such as impaired movement, seen in Parkinson's disease. He also showed that both the chemical deficiency and clinical symptoms can be reversed — at least temporarily — by L-DOPA, a dopamine precursor that is converted into the neurotransmitter in the brain.

Carlsson's discoveries laid the foundations for the work of the US neuroscientists Paul Greengard and Eric Kandel, who share the prize. Studying dopamine-releasing nerve cells, Greengard unravelled the cascade of molecular events needed for a signal to pass across the synapses, the junctions between nerve cells. Kandel showed that changes in synaptic function are essential for learning and memory.

In a series of experiments in the 1950s, Carlsson showed that dopamine was concentrated in parts of the brain involved in movement control. He noted that reserpine, a natural alkaloid used at that time to treat schizophrenia, depleted dopamine stores in the presynaptic neurons. Rabbits treated with reserpine became incapable of voluntary movement, but they recovered when they were given L-DOPA, which compensated for the depleted dopamine.

Carlsson realized that the 'frozen' rabbits were similar to patients with severe Parkinson's, and within a few years L-DOPA was in clinical use. Carlsson also showed that drugs to treat schizophrenia work by blocking dopamine receptors on the surface of postsynaptic nerve cells, stopping the signal from



Nervous Nobels: Kandel (left), Greengard (middle) and Carlsson.

being passed on. Overall, his work has revealed how important dopamine balance is in the brain: too much results in psychosis, too little causes motor disorders.

Greengard, now at Rockefeller University in New York, acknowledges his debt to Carlsson. "One of the reasons I started working on dopamine transmission was because Carlsson had shown the role of dopamine in schizophrenia and shown that antischizophrenic drugs work by disrupting dopamine signalling," he told *Nature*, shortly after learning of his award.

In the 1960s, Greengard began to focus on what happens after dopamine binds to receptors on the surface of postsynaptic neurons. It was known that when some hormones bind to their receptors there is an increase in the level of the second messenger, cyclic AMP. This activates enzymes known as kinases, which add phosphate groups to various proteins, modifying their functions.

"Greengard had the vision and courage to take these concepts of second-messenger signalling processes to the brain — a very tough playground," says Alfred Gilman of the University of Texas Southwestern Medical Center in Dallas, who shared the 1994 Nobel prize for his work on signal transduction within cells.

Over the years, Greengard showed that dopamine, as well as other neurotransmitters, provokes a complicated cascade of phosphorylation and dephosphorylation events. In particular, he found that a protein called DARPP-32 plays a fundamental role

in regulating the phosphorylation states of many of the proteins in dopamine signalling pathways. He also investigated interactions between different pathways. "The more complicated the interaction between different signalling pathways becomes, the more intriguing it all becomes," Greengard says.

Eric Kandel of Columbia University in New York studied the cellular processes involved in learning and memory using the sea slug *Aplysia* as a model. *Aplysia* does not have much to remember — but it does have a reflex to protect its gills. Kandel found that some stimuli amplified this reflex for days or weeks. This 'learning', he showed, arises from an increase in neurotransmitter release at synapses connecting the sensory nerve cells to those that activate the muscles involved in the reflex. This rise is mediated by protein phosphorylation mechanisms similar to those studied by Greengard.

Kandel elucidated the cellular basis of short- and long-term memory in *Aplysia* and has extended the studies to mammals. "He has been a leader in the field and it is appropriate that he is rewarded," says Tim Bliss, head of neurophysiology at the National Institute for Medical Research in London, who in the early 1970s described a molecular mechanism for learning and memory known as long-term potentiation (LTP). "Kandel created the intellectual climate in which LTP could be studied, by showing that changes in synaptic efficiency could lead to changes in behaviour," Bliss adds. ■

www.nobel.se

Fathers of electronic revolution are rewarded

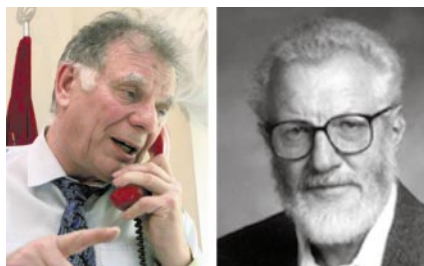
Liesbeth Venema

With the advent of the World-Wide Web and the mobile phone, the information society has truly arrived. This year's Nobel Prize in Physics honours three scientists whose work laid the foundations for modern information and communications technology.

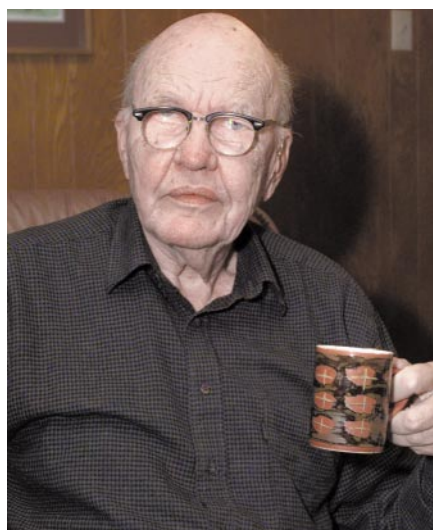
Half of the prize goes to Zhores Alferov of the Ioffe Physico-Technical Institute in St Petersburg, Russia, and Herbert Kroemer at the University of California at Santa Barbara for their work in developing semiconductor heterostructures, which have led to faster transistors and more efficient laser diodes. The other half recognizes Jack Kilby from Texas Instruments in Dallas, co-inventor of the integrated circuit.

Today's electronic devices are both small and fast — we want a mobile phone that slips into our pocket and laptops that fit into our briefcases. At the same time, we expect them to process information almost instantaneously. The three winners have all made major contributions to working out how to get the basic physical principles in small systems working to the advantage of the electronics industry.

The workhorse of electronics is the



Alferov (left) and Kroemer: worked on transistors.



Kilby: co-inventor of the integrated circuit.

transistor, for which William Shockley, John Bardeen and Walter Brattain won the physics Nobel in 1956. But in the 1960s, Alferov and Kroemer started working on faster transistors, made from structures composed of thin layers of semiconducting materials. These layers can each be designed to have specific electronic properties, and it is the subtle interactions between the layers that give heterostructures their advantage over bulk silicon.

One major challenge was to have the layers perfectly matched to each other with atomically smooth boundaries. Alferov and Kroemer showed that transistors built in this way can switch very quickly. They are now in widespread use, for example in the base stations of mobile phone networks. The heterostructures can also be used as small light

sources — laser diodes — examples of which are found in CD players or bar-code readers.

Semiconductor heterostructures have also turned out to be of tremendous importance for fundamental physics. The physics Nobel prizes in 1985 and 1998, for the discovery of the quantum Hall effect and the fractional quantum Hall effect — unanticipated quantum phenomena involving electrons confined within two dimensions — could not have happened without the development of semiconductor heterostructures.

Kilby's main contribution was in miniaturization, as his integrated circuits removed the need for bulky wiring to connect individual transistors. Starting in the late 1950s, Kilby at Texas Instruments and Robert Noyce of Fairchild Semiconductor Corp. in California showed separately that it is possible to etch integrated circuits into thin wafers of silicon. Noyce died in 1990, but Kilby has lived to see his invention become almost ubiquitous.

Laurence Eaves, who works on semiconductor devices at the University of Nottingham, says that researchers in the field will be delighted with the award. "It's a recognition of the basic physics and materials science that has led to the electronic revolution," he says.

In Russia, Alferov's award may be especially important, providing a valuable boost for a scientific community that has suffered immense financial hardship. Russian researchers hope the publicity will help win improved investment in basic science. When *Nature* called Alferov's St Petersburg's office shortly after the announcement, the publicity machine was already in full flow. "He has been seized by the mass media," an assistant explained.

Plastics that conduct win inventors chemistry prize

David Adam

The Nobel Prize for Chemistry has been awarded to Alan Heeger, a physicist at the University of California at Santa Barbara, and two chemists, Alan MacDiarmid of the University of Pennsylvania, Philadelphia, and Hideki Shirakawa of the University of Tsukuba, for their roles in the development of electrically conductive polymers.

The three winners established that polymer plastics can be made to conduct electricity if alternating single and double bonds link their carbon atoms, and electrons are either removed through oxidation or introduced through reduction. The extra electrons or corresponding 'holes' can then move along the 'doped' molecule, making the

conjugated polymer conduct electricity almost as well as a metal.

Some 25 years after the initial discovery, conducting organic polymers — effectively acting as semiconductors — are now being developed for applications ranging from light-emitting diodes in electronic displays to cheap replacements for the silicon chip (see *Nature* 407, 442–444; 2000).

Shirakawa first stumbled across a new method of making *trans*-polyacetylene films in the early 1970s — a silvery film appeared when 1,000 times too high a concentration of catalyst was added by mistake. MacDiarmid and Heeger were also experimenting with a metallic-looking film of the inorganic polymer sulphur nitride. After the three met, they jointly discovered

that iodine doping boosted the electrical conductivity of *trans*-polyacetylene 10 million times, and published the results in a seminal 1977 paper (*J. Chem. Soc. Chem. Commun.* 579; 1977).

Chemists and physicists have taken the field further since then — discovering that organic polymers can be induced to light up, for instance. Layers of these conducting polymers could help to create flat television screens or luminous traffic signs.

"I'm surprised, but I think it is wonderful," says Lewis Rothberg, a researcher into conjugated polymers at the University of Rochester, New York. "A lot of things that started with their discovery are not only of tremendous scientific interest but will also have a great technological impact."

Gore and Bush back rise in science spending

Colin Macilwain, Washington

Both of the main candidates in this year's tightly contested US presidential election are promising to boost research funding — especially at the National Institutes of Health (NIH) — and maintain the existing structure of research agencies, according to statements from their campaign staff.

The similar approaches to science and technology policy taken by vice-president Al Gore, the Democrat nominee, and George W. Bush, the Republican, contrast with some previous campaigns. In 1992, for example, Democrat challenger Bill Clinton pledged to revive US industry with billions of dollars worth of new technology programmes.

This time, the Bush campaign is pledging an extra \$20 billion over five years for military research and development. But it is not clear how much of this would be spent on science, or even on applied science: four-fifths of current military R&D spending goes on weapons development and testing.

The similar approaches were clear at a debate last week between representatives of the Gore and Bush camps, organized by the Washington Science Policy Alliance.

But the debate did reveal sharp differences on areas of policy influenced by scientific knowledge, such as global warming and ballistic missile defence. Bush promises to deploy national missile defence, whereas Gore says he will consider its effectiveness and the views of US allies before deployment. Gore will seek Senate ratification for the Kyoto Protocol on climate change; Bush wants to renegotiate it.

In addition, Bush promises to fight to ensure that US farmers' genetically modified crops are allowed to enter the European Union. David Beier, Gore's chief domestic policy advisor and a former executive with biotechnology firm Genentech, says that "we need to be very careful" about dealing with agricultural biotechnology.

Both candidates are committed to doubling the NIH's budget from its 1998 level of \$13.5 billion to \$27 billion by 2003 or soon after. Both have also pledged more money for the National Science Foundation, with Gore saying that he hopes to double its budget in five years.

Former Republican congressman Bob Walker, speaking for the Bush campaign at the debate, said that cuts in military research under the Clinton administration "had crippled a lot of our universities". But Beier argued that the vice-president's interest in science and technology was "probably greater than that of any presidential candidate in history".

Neither candidate has drawn up detailed proposals for science and technology,



Similar approach: Bush (right) and Gore both promise to boost science funding if elected.

although both campaigns summarized their plans in questionnaires published in *Physics Today* this month and *Science* this week.

Hopes that Bush would devote a speech to science and technology are receding, although Gore may do so shortly, possibly on a visit to Michigan.

Science lobbyists, disappointed at the lack of science and technology in the campaigns, note that the tightness of the race — and the candidates' need to focus relentlessly on hot-button issues such as healthcare for the elderly — has diverted them from speeches or policy papers on less mainstream topics.

Also, Gore is expected to win easily in science-rich states such as California, Massa-

chusetts and Maryland. Of large states where the election could go either way, only Michigan has a major stake in research and development.

The candidates have "much greater differences on other stuff" than on research policy, says Al Teich, head of science and technology policy at the American Association for the Advancement of Science, who helped to organize the debate.

In the *Science* questionnaire, Gore says: "My favourite subject at school was science." Apparently impressed, Nobel laureates Murray Gell-Mann and Harold Varmus are circulating a letter urging scientists and engineers to support the Gore campaign. ■

Project offers free mouse sequence

Paul Smaglik, Philadelphia
& Alison Abbott, Munich

Researchers will get a free version of the mouse genome about two years earlier than planned, thanks to a public-private collaboration announced last week. A consortium will pump \$58 million into the project — enough to sequence the organism three times over by April.

Celera Genomics, of Rockville, Maryland, will finish its mouse project next month, but the information will be available only to subscribers. In supporting the public project, rather than paying to use Celera's databases, the consortium's biggest pharmaceutical sponsors, Merck and SmithKline Beecham, have committed themselves to making what their spokespersons call "precompetitive information available to all".

The companies have each given \$6.5

million to the public effort. Other sponsors include Affymetrix, the US National Institutes of Health and Britain's Wellcome Trust.

In announcing the consortium last week at a meeting of the American Society of Human Genetics, Francis Collins, director of the National Human Genome Research Institute, emphasized that the project is not competing with Celera, in contrast to the portrayal of their sequencing efforts of the human genome. "This is not a race," he said.

The mouse sequence will make it easier to understand the human genome, as the two share many genes — although not the repetitive stretches of 'junk' DNA that make up a significant part of each.

The two mouse projects use different strategies. Celera is sequencing three different strains of mouse once each, whereas the public effort is sequencing the common 'black six' strain three times.



Mighty mouse: knowing the murine genome will help in understanding human genes.

► Celera's effort will allow subscribers to detect subtle differences between strains; the public project will give a more complete view of one strain.

But Celera president Craig Venter sees duplication, not difference. The consortium's effort is a "waste of public money", he says. "It would make more sense for scientists to pay for Celera licences than to pay for the genome to be sequenced again."

Roger Schultz, assistant professor of human growth and development at the University of Texas Southwestern Medical Center, sees merit in the public approach, which aims to sequence both mouse and human genomes several times more than Celera intends.

"Personally, I'm more interested in the high-quality human product, although a good mouse project can help you find highly conserved regions, and therefore genes," he says. The Texas centre is one Celera's several academic subscribers.

Celera subscribers will get the first view of the mouse, as the company expects to finish sequencing next month. If the public project finishes its first phase in March, as planned, the data could be largely assembled by the end of next year. Plans announced by the Human Genome Project last autumn called for a draft of the mouse by the end of 2003, to be fully completed by the end of 2005.

The Whitehead Institute at the Massachusetts Institute of Technology, Washington University in St Louis, and the British Sanger Centre near Cambridge will do the bulk of the public project's shotgun sequencing. Washington University is halfway through building a map of the mouse that will help in assembling a rough draft from the mouse shotgun data. ■

► http://www.nhgri.nih.gov/NEWS/MouseGenes/mouse_release.html

Anger as Spain boosts R&D figures with defence money

Xavier Bosch, Barcelona

Spanish scientists are expressing anger that the government is exaggerating its generosity towards research by including a number of military projects in the national research and development (R&D) budget.

Late last month, the Spanish government announced an 11.3% increase in next year's state spending on R&D, following agreement by the cabinet to boost R&D spending by Ptas 572 billion (US\$3 billion).

According to a report accompanying the budget figures, basic and applied research managed by the Higher Council of Scientific Research (CSIC) and the Ministry of Health will receive a 7.6% increase to Ptas 112 billion. But a significantly larger increase of 12.6% will go to the general categories of 'technological R&D' and 'information society R&D'.

The government report says this spending is intended to boost "competitive technological research as well as to finance the development of the information society". But Ramón Marimon, secretary of state for scientific policy at the Ministry of Science and Technology (MST), admits that the budget for 'technological R&D' — the largest item in the whole R&D budget at about Ptas 300 billions — is devoted to technological projects for military use.

According to information on the website of the former Ministry of Industry and Energy, which has been replaced by the MST, the minister has allocated Ptas 500 billion since 1997 to the Ministry of Defence for the construction of tanks, frigates and fighter planes. These funds have been officially listed as 'R&D' funds, and their military nature has not been explicitly stated.

This military spending may explain the impressive growth of the nation's R&D budget from Ptas 207 billion in 1995 to 507 billion in 2000, which, according to the budget report, supposes "a cumulative annual mean rate [of R&D investment] of 19.6%".

"Not only is the government trying to disguise the true situation, but a significant amount of the R&D budget is spent on military research, while the money set aside for [civilian] research remains scarce," says Óscar Fornas, a senior biomedical researcher at the University of Barcelona.

The MST has admitted to increase the proportion of gross

LATEST SPANISH MILITARY CAMOUFLAGE UNVEILED!



national product spent on research from 0.9% to 2% by 2003, in line with other advanced countries.

But if this figure is reached by devoting a disproportionate amount of funds to military research, then "something is very wrong", says Jordi Camí, professor of pharmacology at the University Pompeu Fabra in Barcelona.

Mariano Esteban Rodriguez, director of the National Centre of Biotechnology in Madrid, says scientists "must know how much money there really is for research", and adds that "transparency is essential to know how much money we can spend".

Luis Rull, professor of physics at the University of Seville, says the situation was the same last year, when it was criticized by the Spanish Association for the Advancement of Science and Technology.

Spaniard Juan Manfredi, professor of mathematics at the University of Pittsburgh, says the way the Spanish budget is presented is significantly different from the situation in the United States, where a clear distinction is made between civilian and military spendings on research.

He adds that a line in the Spanish budget described as military R&D ('research and studies of armed forces') should not be included in the research budget, as most of the money "represents the adaptation and fitting of existing technologies".

While admitting the substantial spending on military projects, Marimon points out that at least 2,000 new research posts will be created in the public sector.

He has also announced that companies involved in R&D will no longer pay value added tax (VAT) on their activities. But he admits that it would be "reasonable" to increase the amount of transparency in the Spanish budget in the future, and promises to try to achieve this. ■



Marimon admits to military spending.

© 2000 Macmillan Magazines Ltd

Astrometry mission wins German approval

Alison Abbott, Munich

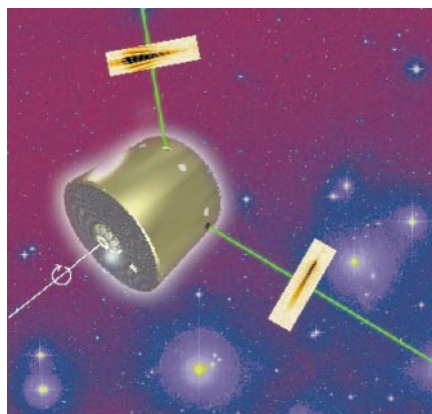
The German Space Agency DLR last week selected DIVA (German Interferometry for Multichannel Photometry and Astrometry), an astrometry mission, as Germany's next small space science mission.

DIVA had been given top ranking by the DLR's ad hoc committee for space research. As a result, it now seems unlikely that there will be a second chance for ABRIXAS, the small German X-ray satellite that was lost last year, two days after launch, when its batteries failed (see *Nature* **399**, 93; 1999).

A proposal for a new mission, ABRIXAS II, was ranked third behind MEGA (Medium-Energy Gamma Ray Astronomy), a gamma-ray astronomy mission.

DIVA will bridge what would have been a 20-year gap in astrometry missions between the switching off of the European Space Agency's mission Hipparcos in 1993 and the agency's sixth cornerstone mission GAIA, scheduled for launch in 2012 (see *Nature* **407**, 277; 2000).

DIVA will measure the positions and motions of about 35 million stars in the Milky Way with high precision, and stars in other galaxies with lower precision. It will view 300 times more objects than Hipparcos, with a fivefold greater precision. GAIA will eventually measure 30 times as many objects as DIVA, with 20 times DIVA's precision.



Map maker: DIVA will plot the positions and movements of millions of stars.

High-precision astrometry missions are important because a precise estimate of the mass of the visible Milky Way and the motions of its stars will help to characterize the amount and nature of the dark matter that astronomers suspect make up nine-tenths of its mass.

DIVA will give European scientists the chance to carry the Hipparcos baton until GAIA arrives on the scene, says Siegfried Röser, DIVA's principal investigator and a senior scientist at the Astronomisches Rechen Institute in Heidelberg. "There is still enough Hipparcos expertise to train the

next generation of astrometers on DIVA, so they will be able to pick up the scientific reins of GAIA."

Extra money will have to be found: DIVA will cost DM100 million (US\$45 million), twice the amount set aside by DLR for its next small mission. Röser is optimistic that three German states with interests in aerospace — Baden-Württemberg, Thuringen and Bremen — will come up with the balance.

DIVA has also been helped by the Klaus Tschira Foundation, set up by one of the founders of the Heidelberg-based software company SAP. The foundation has donated DM1 million to DIVA's data analysis and sponsored an opera to raise money and promote private donations.

A further DM80,000 has been raised in this way. This is "very important for us because it demonstrates public interest in supporting science directly", says Röser.

Günther Hasinger, incoming director of the Max Planck Institute for Extraterrestrial Physics in Garching and a joint principal investigator of ABRIXAS II, says that DIVA's selection was fair. "It's hard to swallow," he admits, "but the community felt that X-ray astronomers were well served by ESA's XMM and NASA's Chandra."

He adds that ABRIXAS scientists are seeking alternative international partners to help finance the DM50 million project. ■

Ig Nobel glory for levitating frogs and collapsing toilets

Steve Nadis, Boston

Intelligence — and the lack of it — was the theme of this year's Ig Nobel prize ceremony, which celebrated its tenth anniversary at Harvard University last week.

Once a year, scientists gather to poke fun at themselves and others, hopefully without hurting too many feelings or careers. This time, ten prizes were awarded to winners from eight countries.

American researchers David Dunning and Justin Kreuger took the psychology prize for their 1999 paper, "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments".

Canadian biologist Richard Wassersug won the biology prize for his study, "On the palatability of some dry-season tadpoles from Costa Rica". On claiming the award, Wassersug explained that the tadpoles were "neither dry nor seasoned".

The physics prize went to Andre Geim of the Netherlands and Sir Michael Berry of England for using magnets to levitate both a frog and sumo wrestler. Magnets also

featured in the medicine prize, awarded to Willibrord Weijmar Schultz, Pek van Anel and Eduard Mooyaart of the Netherlands for their MRI scans of male and female genitals in the act of coupling.

Scottish researchers Jonathan Wyatt, Gordon McNaughton and William Tullet captured the public health prize for their alarming report, "The collapse of toilets in Glasgow". Wyatt told the audience that the research had been dismissed for years "as a flash in the pan".

Three genuine Nobel laureates — Charles Clements (peace, 1997), Dudley Herschbach (chemistry, 1986), and Richard Roberts (physiology or medicine, 1993) — participated in the world premiere of *The Brain Food Opera*.

They took the stage to salute the King and Queen of Swedish Meatballs, asked the audience to stand for a "Moment of Science", and judged the "Great Intelligence Debate", in which participants squared off for 30 seconds, talking simultaneously.

The event ended with words of encouragement from master of ceremonies



Inconvenient: Ig Nobel demonstrators remember to wear appropriate safety equipment.

Marc Abrahams, editor of the *Annals of Improbable Research*: "If you didn't win an Ig Nobel prize tonight, and especially if you did, better luck next year." ■

♦ <http://www.improbable.com/ig/ig-top.html>

Plans drawn up for xenotransplantation watchdog

Paris A global system to monitor adverse events in clinical trials of xenotransplantation is to be set up under the auspices of the Organisation for Economic Co-operation and Development (OECD) and the World Health Organization (WHO). The scheme is intended both to monitor the safety of trials of the controversial technology and to study the risk that pathogens might spread from animals to humans and create new pandemics.

The rough outlines of the proposed system were discussed last week at a meeting of around 70 experts, regulators, industrialists and scientists organized in Paris by the OECD, the WHO, the Canadian ministry of health (Health Canada), which has taken a particular interest in the issue, and other bodies.

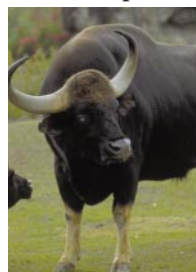
Elettra Ronchi, coordinator of health and biotechnology activities at the OECD, emphasized that the plans were “not an endorsement of xenotransplantation” but an acknowledgement of the reality that trials are already underway. At the very least, the system would comprise a loose international network of experts to provide a rapid response to any worrying adverse events.

Delegates at the meeting said that the challenges of monitoring xenotransplantation required a “new paradigm”.

Cloned gaur being carried by surrogate mother cow

London The first cloned animal from an endangered species is due to be born next month, according to US researchers. A skin cell from a dead male gaur — a south Asian ox-like animal — has been fused with a cow egg cell from which the DNA had been removed. The surrogate mother cow now carries a developing gaur calf. Some scientists think that other endangered species could be cloned and extinct animals revived.

The researchers, from Advanced Cell Technology (ACT) in Worcester, Massachusetts, have named the cloned gaur Noah, and are turning their attention to the bucardo, a Spanish mountain goat. They have



also considered cloning giant pandas using black bears as surrogates, although some cloning experts warn that applying such procedures to even very closely related species can be problematic.

Human-rights workers win alternative Nobels

Stockholm Four alternative Nobel prizes were awarded last week in Stockholm to environmentalists and workers for human rights. The prizes, worth DM460,000 (US\$205,000), are supported by the Right Livelihood Award foundation, created in 1980 by Jakob von Uexküll from the proceeds of the sale of his stamp collection.

This year's awards went to Birsal Lemke for her campaign against the use of cyanide in the extraction of gold from ore in Turkey; Tewolde Berhan Gebre Egziabher, for representing the interests of the Third World at the 1999 biodiversity conference in Colombia; the US agricultural geneticist Wes Jackson for his research on ecological agriculture; and the Indonesian human rights lawyer Munir.

♦ <http://www.rightlivelihood.se>.

US Congress increases allocation of H1-B visas

Washington The US Congress has passed legislation to increase the number of special visas for highly skilled temporary workers to 195,000 for each of the next three years. Last year's quota of 115,000 H-1B visas — available to immigrants with special skills such as

computer programmers, engineers and scientific researchers — was exhausted within six months, leaving universities and employers unable to bring in staff from abroad.

In the past, efforts to raise the number of H-1Bs has met stiff resistance from trade unions and opponents of immigration (see *Nature* 404, 800; 2000). But unemployment is low, and high-technology industry has increasing political influence through hefty contributions to both parties, so the new legislation was passed almost unanimously by Congress.

Europe to target *Arabidopsis* genes

Paris The European Commission is funding two consortia to do post-genomics research on the plant *Arabidopsis*, the genome of which is expected to be published shortly. The Exon Trapping Insert Consortium (EXOTIC) and the Regulatory Gene Initiative in *Arabidopsis* (REGIA) will share 11 million euros (US\$9.6 million). Both collaborations include research groups from ten European countries.

REGIA aims to study virtually all *Arabidopsis* transcription factors; EXOTIC will focus on patterns in gene expression, mirroring commercial work under way in the United States. The three-year projects will provide a “backbone to make steps in

identifying gene function”, says EXOTIC project manager Jonathan Clarke at the John Innes Centre, Norwich.

Parallel home computers to model protein folding

Stanford Vijay Pande of Stanford University's chemistry department is hoping to do for protein folding what the SETI@home screensaver has done for the search for extraterrestrials. By exploiting idle processor time on computers across the Internet, Folding@home (<http://foldingathome.stanford.edu>) hopes to achieve massive distributed computing to model protein folding. SETI@home is now the world's largest supercomputer.

The challenge of modelling how proteins fold is already the focus of a \$100 million research effort by IBM to build a petaflop supercomputer known as ‘Blue Gene’. So far, Pande has signed up over 5,000 computers and hopes to attract 100,000.

Korea plans its own Silicon Valley

Seoul Korea is to have its own Silicon Valley, according to a statement made late last month by President Kim Dae-Jung. Companies moving to, or starting up in, Taedok Valley, southeast of Seoul, will get tax exemptions,

government subsidies, funds from the city of Taejeon and the Ministry of Science and Technology, governmental infrastructure investment, and exemptions from military service for their employees.

It is hoped that by building on Taedok Science Town, home to some of Korea's top basic research organizations, Taedok Valley will nurture rapid growth in information technology and biotechnology. The move was seen as an attempt to make Korea's economy less centred on Seoul and less dependent on a few large corporations.

Britain appoints King as chief scientific adviser

London David King, head of the chemistry department at the University of Cambridge and master of Downing College, has been appointed chief scientific adviser to the British government. He succeeds Sir Robert May, whose five-year term of office ended last month. King was tipped as favourite for the post two months ago (see *Nature* 406, 667; 2000).

Born in South Africa, King moved from the University of Witwatersrand to Imperial College London, followed by the universities of East Anglia and Liverpool. In a statement, Prime Minister Tony Blair said that King was “admirably suited to his new role”.

Science for art's sake

In several labs around Boston, the techniques of genetic and tissue engineering are being used in the name of art. Steve Nadis asks the artists and scientists involved what they gain from this fusion of high culture and cell culture.

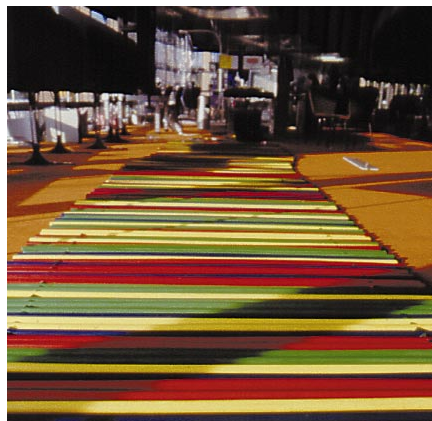
Visitors entering one of the tents set up at last month's Ars Electronica, an international art festival in Linz, Austria, were greeted by an unusual sight. On display, behind the locked glass doors of a refrigerator, sat petri dishes containing *Escherichia coli* bacteria. But these were no ordinary *E. coli*. Their genes included artistically engineered elements of DNA. In one work, called *Microvenus*, the DNA bases constituted a code that, when properly deciphered, provided a symbolic rendering of female genitalia. In another dish, the DNA-encoded message read: "I am the riddle of life. Know me and you will know yourself."

These artistic DNA molecules are the work of sculptor Joe Davis, who for the past decade has been a research affiliate in the laboratory of Alexander Rich, a structural biologist at the Massachusetts Institute of Technology (MIT). Davis, one of a handful of artists using the tools of cell and molecular biology in their work, is the leader of an emerging 'new Boston school' of bioartists.

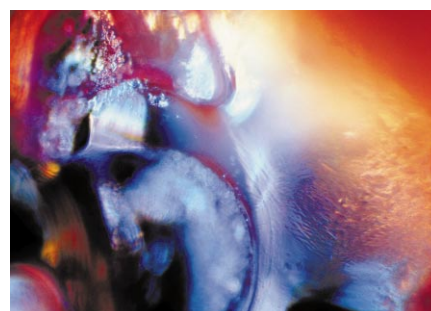
The relationship between science and art has been explored extensively in recent years. But most ventures in this area have involved artists working in traditional media, inspired by the theories or practice of science. Davis, on the other hand, directly engages the techniques of science in his art. "Drawing a picture of something you don't understand is just not good enough," he says. "Art is about communicating. How can you convey something you don't have a clue about?"

Over the years, Davis has devised a series of ambitious projects. In 1992, in an attempt to find a new way of communicating with extraterrestrials, Davis started to search for





Art attack: muscle tissue grown over a hydrogel in the shape of a spearhead (opposite) by Oron Catts, Ionat Zurr and Guy Ben-Ary. Clockwise from top left, Joe Davis, father of the Boston bioart movement; a representation of the DNA code in Davis's *Riddle of Life* bacteria; Adam Zaretsky, one of Davis's recruits, in the lab; muscle grown over a hydrogel by Catts and Zurr; and Davis's audio microscope, which converts the movements of microorganisms into sound.



bacterial spores that might be resilient enough to carry his DNA codes into space. Working with Stefan Wöhl — a former Rich-lab biologist now at the Hans-Knoll Institute in Jena, Germany — and Michael Shia of Boston University, he collected samples of cooling water from MIT's nuclear reactor. The team hoped to find microorganisms that could withstand the radiation and temperature extremes of the reactor's environment. They identified six different bacterial strains in the reactor water.

Another project, called *New Wave Ruby Falls*, involved a plan to put an electron gun on a space shuttle to generate artificial Northern Lights that would be visible from Earth. And in 1993, Davis won another shuttle slot for a project called *Norton Rings*, named after Ed Norton, a fictional sewer worker on the popular US TV show *The Honeymooners*. Davis claims the Earth is orbited by rings of urine and faeces expelled from spacecraft over the years. He plans to install 'fishing gear' in the shuttle's cargo bay to trawl the rings for microorganisms. But all of Davis's space-based projects are still waiting to fly due to a lack of funds.

Still, Davis has managed to secure the intermittent help and interest of dozens of researchers from various institutions including MIT, Harvard University and Boston University. "Joe is able to attract so many scientists because his ideas are fascinating," says Shuguang Zhang, associate director of MIT's Center for Biomedical Engineering. "He's not confined to the normal dogmas of biology. All of his thinking is outside the box."

One of the latest inventions to spring from Davis's fertile mind — the audio microscope

— fits that mould. Lasers are beamed onto a conventional microscope slide holding microorganisms. As they move around they alter the light reflected from the slide. These changes are converted into sound, which is broadcast through speakers. At the same time, the microscopic image is projected onto a video monitor, so viewers can correlate a microbe's motion with its characteristic 'sound'. Using a frequency analyser, the sound's spectrum is determined and is also displayed — showing, for instance, that all paramecia share a similar acoustic signature.

A bacterial cell darting across the screen "sounds like a herd of buffalo," Davis says. "It's exciting. Every time we get our hands on a new organism, we're the first people in the world to hear these sounds. Someday I'd like to incorporate them in an opera or symphony."

Designs for life

Davis's enthusiasm for his work is infectious, but do the scientists who rub elbows with him gain any real advantage from the association? In terms of conventional measures of scientific success, the benefits are hard to quantify. "I'm not aware of any spillover effect of his work into science at our lab," Rich says. "But Joe has lots of unconventional imagination and it's fun to have somebody like that around. He adds another level of interest to this place."

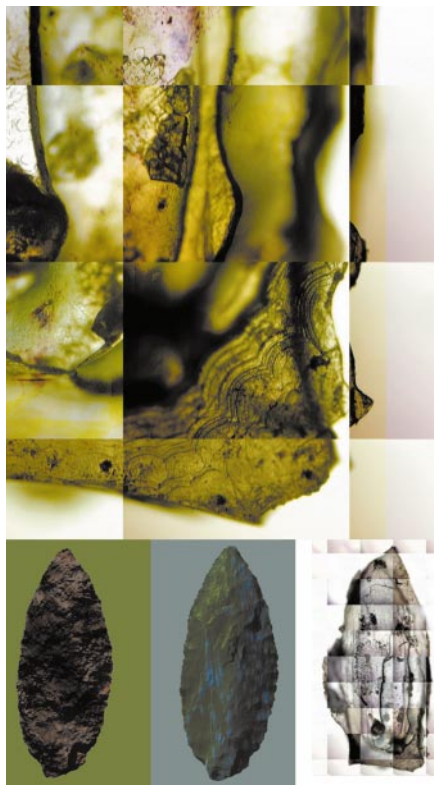
Biologists who have teamed up with Davis find it refreshing to exchange ideas with someone whose thinking deviates from the narrow channels encouraged by traditional scientific training. "Joe takes on projects that are well beyond the scope of what most scientists would consider," says Alan Herbert, a

biologist at the Boston University School of Medicine who worked alongside Davis in Rich's lab for about a decade. "He stretches the boundaries of what is possible."

And aspects of Davis's thinking can rub off on his collaborators. After cooperating with Davis on several projects, Harvard Medical School biologist Dana Boyd says he began designing vectors for gene transfer in the "most elegant fashion, as if they were pieces of craftsmanship, rather than doing it in the usual fastest and cheapest way".

But there are detractors. One biologist in Rich's lab, Yang Kim, considers Davis a "disruptive element", whose bench space would be better used by a scientist. "He acts like this is a playground," Kim says. Another MIT biologist, who asked not to be named, considers it odd that Davis's artwork "is not evaluated with the same scrutiny as the scientific output of the postdocs here". He admires Davis's sculptures, which are on display at MIT and elsewhere in Cambridge, and likes him personally. But in the competitive world of molecular biology, he cannot justify granting precious lab space to an artist. "If every lab had a Joe Davis, the university would probably raise a stink about overhead costs," he says.

There is little risk of that, says Rich. "There aren't that many Joe Davises running around. He's a rare bird, not a widespread phenomenon." Yet Davis is actively encouraging other bioartists to move to Boston. At the moment he is working with artist Katie Egan on various projects including the audio microscope, while trying to expand his circle of colleagues. Plans for the field's first meeting, which is to be held at MIT, are in the works.



Would you let this man loose in your lab? Zaretsky sets out his art agenda.

But the sparse funding for bioart means that recruitment has been difficult. Sympathetic scientists such as Rich might help out by offering a lab berth, but they cannot offer financial assistance. Davis admits to being broke, and bemoans an art establishment that has yet to embrace new, biological forms of expression. "The current funding system is set up to support bygone ways of art," he says.

Prevailing attitudes within the scientific establishment are also imposing limits on bioart's development. For example, Davis is trying to find a laboratory position for Graham Smith, chief technology officer with Telbotics, a robotics company in Toronto. Smith and Davis plan to design robots controlled by individual cells. The Canada Council for the Arts has even promised to fund Smith's work if Davis can get him installed in a Boston lab. But one local biologist, who initially was intrigued with the idea of taking on an artist, eventually declined. "As a junior faculty member, that's not the sort of thing that will help me get funded or tenured," he says. "When I'm as established as Alex Rich, it might be a possibility."

Indeed, for the time being, it seems that only senior scientists such as Rich — who was awarded the National Medal of Science in 1995 for his work on the structure of nucleic acids — can afford to ignore the attitudes of colleagues and bring artists into their labs.

Despite the obstacles, Davis is succeeding in enlisting artists to his Boston movement. Another recent recruit is Adam Zaretsky. As a graduate student at the Art Institute of Chicago, Zaretsky first heard Davis speak there in 1998 and was "blown away". After earning his

masters degree in fine art, he came to MIT last year. Once Davis was satisfied that the young artist had fully embraced the bioart philosophy, he helped Zaretsky obtain a position in the lab of Arnold Demain, an MIT biologist who had once entertained dreams of becoming an artist himself. "Maybe Adam's here because I was a frustrated artist early in my life," Demain says.

Sound science?

Zaretsky received a two-year unpaid appointment to pursue a project that Demain describes as "too wild for any postdoc to take on". The idea, which is an offshoot of Davis's audio microscope, is to see whether sound or music can influence the behaviour of an *E. coli* strain engineered to produce the antibiotic microcin B17. "If we could stimulate or inhibit antibiotic production, that would be fantastic," Demain says, acknowledging that this project is an incredible longshot that would be viewed by many scientists as a frivolous endeavour.

Applying the scientific method has not come easily to Zaretsky. "The hardest part is learning to be meticulous, when as an artist you're supposed to be just the opposite," he says. In one experiment, Zaretsky played the music of Engelbert Humperdinck to his bacterial subjects for two days straight. He thinks he just might have observed a response in the cells — the details of which are being kept quiet, pending verification. If Zaretsky can reproduce these results, which he jokingly calls 'the Humperdinck effect', Demain will get one of his postdocs to try to replicate them. "We have to worry about artefacts," Demain says. "With something like this, you don't want to go into the literature unless you're sure you're right."

Growth industry

Across the Charles River, two other members of Davis's Boston school are spending a year as research fellows in Joseph Vacanti's tissue engineering lab at the Massachusetts General Hospital. Vacanti's lab is famous

for creating grafts for transplant surgery by growing tissue, such as cartilage and muscle, on polymer scaffolds. But Oron Catts and Ionat Zurr, who were previously based at the University of Western Australia in Perth, use tissue engineering technology to create living, growing sculptures. With their Perth colleague Guy Ben-Ary, Catts and Zurr showed their work at Ars Electronica next to Davis's exhibit.

Catts and Zurr are, so far as Vacanti knows, the first artists-in-residence in the hospital's 189-year history. "I see part of their role as trying to reduce the barriers between the science world and the world that artists represent," says Boris Nasser, a researcher in Vacanti's lab. "When I tell my colleagues we have artists working here, first they are sceptical, then curious, and then enthusiastic," adds Vacanti. But, like Rich, Vacanti approaches the collaboration safe in the knowledge that his lab is already recognized as a world leader in its field. Consequently, he does not have to worry about whether the scepticism voiced by his colleagues will damage his career prospects.

Vacanti relishes the unpredictability of Catts and Zurr's work, hoping that their presence will add "a new dimension" to his lab. "They can illustrate the beauty that we gloss over while doing quantitative research," he says.

Catts agrees, but says that from the artist's point of view, there is a more fundamental motivation for linking up with leading scientists. "Artists need to comment on the world around them," says Catts. That world, he adds, should include science and technology — and biological research in particular — which are emerging as key driving forces behind the development of twenty-first-century society. "We can't go on painting landscapes forever, as if nothing has changed," he says.

Steve Nadis is a writer in Cambridge, Massachusetts.

Ars Electronica ▶ <http://www.aec.at/festival2000>

Tissue Culture and Art Project ▶ <http://www.tca.uwa.edu.au>

Poorly conducted (or reported) animal tests put humans at risk

Sir—Animal clinical research paradigms are part of the mechanism for assessing the risks of new drugs to humans. To extrapolate the results to humans requires that animal experiments are designed adequately and conducted properly, with the methods fully reported. But these goals are often not met, either because collection of equivalent clinical parameters is held to be too difficult or time-consuming, or because of an effort to oversimplify complex biological systems in animals that are not identical or useful to the human paradigm^{1–3}. This problem needs rectifying.

The sort of information that should be recorded and reported includes means and standard deviations for weight, electrolytes and glucose and the exact time of day of key treatments (including weekends). This would reveal whether such variables were kept constant throughout the experiment, and would show whether failures were due to failure of treatment or of support care.

Another area requiring more detailed reporting is that of pain detection and alleviation, which is not evident in publications of studies supported by the US Public Health Service (PHS). The PHS and the US Department of Agriculture require alleviation of pain in most animals, but this information isn't appearing in the methods sections of publications.

In order to provide information about the dynamic nature of the success or failure of treatment, clinical pathological data should be used, from tests such as complete blood count, clinical chemistries and organ profiles, carried out on living animals at specified times during a disease. Pathological data alone (from tests carried out after death, such as liver biopsies) are not adequate.

Omissions in reporting are prevalent. Take for example recent problems in human gene therapy⁴. Primate experiments clearly suggested risk. But in the animal work, the nature of the morbidity in one paper⁵ was described only in terms of pathological data. The reader is informed that appetite in the survivors was normal, but weight changes were not reported. It is not recorded whether the animals were housed in groups (usually required for primate social needs) or singly, leaving the knowledgeable reader wondering whether a cage-mate ate the food. Additionally, mortality was described only by pathological criteria based on liver biopsies, rather than including information from, for example, liver function tests, complete blood count

information, haemodynamic criteria and subjective scores in both survivors and dead animals. This left the events occurring between life and death open to speculation.

In a slightly improved reporting scheme in canine research⁶, when using pericardial adenoviral-mediated vascular endothelial growth factor (VEGF), investigators reported a severe dose-dependent physiological reaction after administration. Substantial information about morbidity and mortality was included, but key aspects about related events were omitted. For example, no information was given about intense electrolyte changes, nausea and occasional thrombocytopenia — seen and treated by the veterinary staff — which could have contributed to the later drug and technical failures of VEGF and adenovirus in human patients⁷.

Scientists and the public need access to correct information about clinical protocols in animals. We need more of the right kind of clinical study design, and more collaborative reporting from inside animal facilities.

Such partnerships between veterinarians and scientists are stuck — they require a change in thinking. Scientists need to stop viewing veterinary care as a burden and see it as a quality assurance for drugs that are destined for human use. As well as allowing a more accurate risk-benefit analysis for human drug trials, introduction of such reporting standards in non-human experiments could also reduce the amount of costly regulation required at the later stages of commercialization of drugs.

Victoria Hampshire

Advanced Veterinary Applications, 7307 Nevis Road, Bethesda, Maryland 20817, USA

1. Scientists Center for Animal Welfare: <http://www.scaw.com/summary.htm>
2. NIH Initiative to Reduce Regulatory Burden. <http://grants.nih.gov/grants/policy/regulatoryburden/index.htm>
3. Foundation for Biomedical Research. <http://www.fbrresearch.org/press-ardf.html>
4. *Washington Post* 12 July (2000).
5. Nunes, F. A., Furth, E. E., Wilson, J. M. & Raper, S. E. *Hum. Gene Ther.* 15, 2515–2526 (1999).
6. Lazarous, D. F. *et al. Cardiovasc. Res.* 2, 294–302 (1999).
7. *USA Today* 3 November (1999).

Why don't creationists use private schools?

Sir — In his recent review of Niles Eldredge's *The Triumph of Evolution and the Failure of Creationism* (*Nature* 406, 935; 2000), Robert W. Cahn speculates on why creationism has surfaced only among Christian fundamentalists in the United States. He says ultraorthodox Jews and fundamentalist Muslims are more concerned with daily ritual, dietary practice and appropriate observance of

holy days, while fundamentalist Christians are wholly focused on belief — causing the current ground swell of activism to curb the teaching of evolution in US schools.

There is no reason to believe that ultraorthodox Jews and fundamentalist Muslims are any less fervent in their beliefs about creation by an all-powerful deity. However, most choose to educate their children in private religious schools where they can control the science curriculum without infringing on the rights of others to learn about evolutionary theory.

Apparently, Christian fundamentalists feel they are entitled to publicly sponsored and funded schools where the teaching conforms to their religious beliefs, even though this is prohibited by the US constitutional separation of church and state. Perhaps they should be encouraged to send their children to private schools, so that other children are no longer deprived of their right to be taught a current, well-validated scientific theory.

Jeffrey M. Marcus*, **Joanne E. Seif†**

**Box 90325, Department of Zoology, Duke University, Durham, North Carolina 27708-0325, USA*

†CB 3225, Department of Religious Studies, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

Survival on the edge: the tube worm's strategy

Sir — In his otherwise excellent Millennium Essay¹, Roel Snieder proposes an analogy between scientific specialists and tube worms found at deep-sea hydrothermal vents. He advocates *Homo universalis*, an interdisciplinary 'team-player' who can communicate effectively with colleagues from other fields.

But although individual tube worms themselves don't move into new (vent) fields when conditions change, it is now becoming apparent that their notable symbioses with chemoautotrophic bacteria, their remarkable physiological adaptations and their impressive reproductive capabilities allow their species to deal quickly with changing conditions by colonizing new opportunities (vents) — and yes, even interbreeding among different colonies, in time².

And a good thing, too, because deep-sea vents at fast-spreading centres are, individually, some of the most unstable habitats on this planet. Perhaps tube worms would make good interdisciplinary scientists, after all.

John D. Rummel

NASA Headquarters, Washington, DC 20546, USA

1. Snieder, R. *Nature* 406, 939 (2000).
2. Van Dover, C. L. *The Ecology of Deep Sea Hydrothermal Vents* (Princeton Univ. Press, 2000).

Activism, scientists and sociobiology

A scientific movement honed by capitalists, Marxists — and the Vietnam War.

Defenders of the Truth: The Battle for Science in the Sociobiology Debate and Beyond

by Ullica Segerstråle

Oxford University Press: 2000. 493 pp.
£20, \$35

David L. Hull

For more than 20 years Ullica Segerstråle has been charting the course of sociobiology, beginning with E. O. Wilson's *Sociobiology* (1975) and Richard Dawkins' *The Selfish Gene* (1976) through to the present-day 'science wars' and evolutionary psychology. Although she is interested in broad sociological and philosophical trends, her exposition here consists mainly in discussions of individual people, their views and their interrelations.

The chief advocates of sociobiology in the United States whom Segerstråle considers are E. O. Wilson, Robert Trivers and Bernard Davis; the chief opponents are Richard Lewontin, Stephen Jay Gould, Richard Levins, Jon Beckwith and Stephen Chorover. In the United Kingdom she emphasizes the work of Richard Dawkins, W. D. Hamilton and John Maynard Smith; Steven Rose and Patrick Bateson are the chief critics. In the

United States Wilson emerged as the father of sociobiology whereas in Britain this role devolved on Richard Dawkins, although quite understandably he preferred to call this movement by a name other than the one co-opted by Wilson.

Segerstråle interviewed all the major figures in these disputes and many of the minor figures too. She also attended meetings, both formal and informal, and read the vast literature that sociobiology generated. Her goal was to understand the factors that influenced the course of this scientific movement.

The controversial thread that runs through her narratives is the nature of science. Science functions within a culture. So, do the various sociocultural contexts influence science? Do capitalist societies necessarily produce capitalist science? According to the traditional 'internalist' view of science, scientists make up their minds on scientific issues primarily through reason, argument and evidence. Other factors may be involved, but they should be weeded out.

Just as sociobiology was emerging, a quite different view of science was becoming influential. According to this view, scientists make their decisions in large measure influenced

by broader social contexts, such as those of economics and class structure.

More specifically, Segerstråle attempts to discover exactly what the views of the biologists she studied were and why they held them. On what basis do the sociobiologists as well as their opponents evaluate sociobiology? For example, the versions of evolutionary theory that sociobiologists extended to behaviour and social structure tended to be very individualistic and competitive. Sociobiologists tend to think that selection occurs only at the lowest levels of organization, a position their critics attribute to their economic leanings: the individual is paramount in free-enterprise economic systems. The Marxist opponents of sociobiology tend to think that selection can occur at higher levels of organization, including groups. In Marxism, groups are more important than individuals. Capitalists view nature as competitive, whereas these Marxist critics tend to view it as being much more cooperative.

As Segerstråle notes, one problem with posing the issue in the way she does is that sociobiology's opponents lived in exactly the same array of societies and subsocieties as their opponents. During their formative years, nearly all of the protagonists in this controversy were raised in competitive, sexist and racist societies. Why did some of them internalize these features of their societies whereas others did not? Was Wilson really a racist, or did his work just exhibit tacit racism? Segerstråle makes no mention of anyone calling Lewontin a racist. How did he avoid picking up this feature of his society?

According to externalists, political leanings influence the scientific views that scientists hold. Lewontin, Levins and Gould are Marxists; hence, their views on evolution should be influenced by their Marxism. But John Maynard Smith was a more active Marxist than any of these people. Yet he held and still holds views on evolution that are at variance with those of other Marxists and in support of such capitalist running dogs as Wilson and Dawkins. If both internal and external factors affect the course of science, these influences are extremely complicated and at times they conflict.

Segerstråle does not just relate what she has read or what her respondents have told her; she evaluates it and passes judgement on it. Looking back over the past quarter-century, she considers one of the gratifying developments to have been that we have a "relative vindication of the sociobiologists unfairly accused at the beginning of the controversy".

To complicate matters further, Segerstråle was engaged in the same sort of activity as her



Sign of the times

Napalm Flag is artist Leon Golub's angry gesture over the Vietnam War, a "symbolic icon of nationhood as a visceral and bloody stain".

Leon Golub: Echoes of the Real by Jon Bird (Reaktion Books, £17.95, \$29; pbk) assesses the history painter's life and work.

book reviews

subjects. She was a scientist studying scientists, a meta-scientist if you will. She had to make decisions about what she thought she was doing. The fact that she spends a lot of time explaining the relevant science implies that she thinks it matters. If it can influence her, it can influence other scientists as well. This problem confronts all students of science. How we study science implies something about what we take science to be.

As Segerstråle sees it, the significant difference between Wilson and Lewontin was in their attitude towards science. Wilson was willing to take chances, to come up with new ideas and to pursue them even if they seemed implausible or overly ambitious. He admits that his early efforts to biologize all of the social sciences, not to mention the humanities, might seem too simplistic. But he says the beginnings of general theories come out of such oversimplifications. Lewontin, in contrast, possibly because he thinks that such things as social class can influence science, holds a hard-nosed attitude to science — new theories must be clearly formulated and backed up with significant amounts of data.

Accurate though her explanation of the differences between Wilson and Lewontin might be, Segerstråle pays insufficient attention to one crucial aspect of the sociopolitical context of the time — the Vietnam War. Many Americans felt helpless during this time. They were faced with a lot of problems, not the least of which was a cruel, stupid war about which there was so little they could do. They could sign petitions, march in protest and burn draft cards, but that was about it. Early in her discussion, Segerstråle remarks that the sociology controversy was not between the left and right. "The actual dividing line went, rather, between a particular type of New Left activist on the one hand and traditional liberals and democrats on the other." The key term is "activist". The battle waged against sociology was part of this activism.

I must also mention the most famous incident of all. In 1978, at a meeting of the American Association for the Advancement of Science, both Segerstråle and I attended a session on sociobiology at which Wilson was to present a paper. As he began his presentation, a dozen or so members of the International Committee Against Racism marched up onto the stage, chanting: "Racist Wilson you can't hide, we charge you with genocide!" A woman then poured water over Wilson's head. How much water is a matter of conjecture. Usually we are told it was a pitcher of water. Segerstråle remembers a jug. I am sure that it was a small paper cup. One bit of evidence that supports my memory of the incident is that Wilson was able to mop up the water with a single handkerchief. Such are the problems of eye-witness reports. ■

David L. Hull is in the Department of Philosophy, Northwestern University, Evanston, Illinois 60608, USA.



The climate of fashion

Artist Lucy Orta's *Refuge Wear. Collective Survival Sac — 2 Persons. With Transformable Rucksack*. From a brief

pictorial overview, *Art & Fashion* by Florence Müller (Thames & Hudson, £12.95).

Dowsing the human volcano

Something New Under the Sun: An Environmental History of the Twentieth-Century World

by J. R. McNeill
Allen Lane/W. W. Norton: 2000. 421/448 pp.
£20/\$29.95

Paul Crutzen

Looking back at mankind's impact on the environment during the twentieth century, with J. R. McNeill, professor of history at Georgetown University, as a guide, one is both impressed and depressed. To take a few examples from this highly informative book: the world population grew by a factor of four to around 6,000 million; the urban population increased 13-fold; industrial output increased 40 times and energy use 16 times. In the twentieth century, according to McNeill's estimate, humans used ten times more energy than during the whole of the

rest of the millennium. Water use increased by a factor of seven, and the methane-producing cattle population paralleled the increase in the human population; on average, one cow per family supplies humans with dairy products and meat. The fish catch grew 35 times; and emissions of carbon dioxide and sulphur dioxide grew by a factor of more than ten.

The Earth's carbon, nitrogen and sulphur cycles are now strongly perturbed by agricultural and industrial activities. The global release of sulphur on the continents as a result of the burning of coal and oil — the "human volcano" — is an order of magnitude larger than all natural inputs combined. The supply of nitrogen to the environment from fertilizer application and from fossil-fuel burning is of a similar magnitude to total global biological nitrogen fixation.

The environmental impacts on the atmosphere are well known: unhealthy air to breathe in industrial cities, acid rain, photochemical smog, and increases in the greenhouse gases carbon dioxide, nitrous oxide

and methane. Even the apparently benign chlorofluorocarbons have turned out to be an environmental menace, delivering chlorine to the stratosphere, where it acts as a catalyst in ozone destruction. The result has been the creation of the 'ozone hole' over the Antarctic, a region previously considered to be least vulnerable to such effects.

Many of the human effects on the environment are long term. The average residence times of chlorofluorocarbons, carbon dioxide and nitrous oxide in the atmosphere are 100 years or more, and some of their effects, such as global warming and the resulting sea-level rise, may linger for "centuries to millennia". The effects of human contaminants such as nitrate and lead on soil and groundwater could last even longer.

McNeill's book contains hundreds of examples of environmental assaults by humans on atmosphere, hydrosphere, pedosphere and biosphere, in all parts of the world, during the last century. All are described clearly and lucidly.

But the book is more than a mere summing up of human damage to the environment; it looks for links between the history of the planet and that of its peoples. McNeill doesn't see only the negative effects of the developments of the twentieth century. The engine, the source of the environmental problems, also contributed to the liberation of some 25% of the world's population from hard labour, opening up for them possibilities for better education and the pleasures of free time. But all this comes at a considerable price, and the big question is whether such developments can be made sustainable and, even more importantly, whether the other 75% of the human population can reach the same standards of living without the total destruction of the Earth's environmental base.

McNeill does not try to give answers and inspires readers to make up their own minds. However, in the final page of his book he gives some important advice: "Modern history written as if the life-support systems of the planet were stable, present only in the background of human affairs, is not only incomplete, but is misleading. Ecology that neglects the complexity of social forces and dynamics of historical change is equally limited. Both history and ecology are, as fields of knowledge go, supremely integrative. They merely need to integrate with one another."

I am very impressed with the book, and strongly recommend it to policy-makers and to historically and environmentally interested readers. It should serve well as teaching material at universities and even high schools. It is an important, beautifully written book. ■

Paul Crutzen is at the Max-Planck-Institute of Chemistry, PF3060, 55020 Mainz, Germany.

Icy displays from before time

Comet Science: The Study of Remnants from the Birth of the Solar System

by Jacques Crovisier & Thérèse Encrenaz
Cambridge University Press: 2000. 173 pp.
£37.50, \$54.95 (hbk); £14.95, \$19.95 (pbk)

Dale P. Cruikshank

As recently as 20 years ago, the most challenging part about getting new observations of a comet was the process of making some sense of them. This was particularly true for the photographic images that characterized a century of comet studies. We could see little more than signs of a comet's motion and the behaviour of its tail and the 'coma' of dust and gas as it approached, and then receded from, the Sun. We were given only vague clues to the composition, size and origin of the intriguing solid lump that lay inside the opaque comatic fog enshrouding the heart of the matter, the nucleus.

But now comet science is plunging ahead with an exuberance born of a chain of lucky chances, both in space and here on the home planet. Some of the dizzying delight of comet specialists occurs in the glare of the public view, as with the spectacular apparitions of comets Hyakutake (1996) and Hale-Bopp (1997). But most of the deepening and honing of our understanding of these icy bodies takes place in the dimly lit world of the astronomer,

who dispatches spacecraft to selected comets, while probing back on Earth with ever larger and more sensitive telescopes.

The story of these events is told in a clear, succinct and authoritative narrative in *Comet Science*, written by two French astronomers who have each contributed mightily to the modern understanding of comets. The authors trace the history of observations of comets and their role in the development of science since the time of the Greeks. But the bulk of the book focuses on the revelations of comet behaviour, composition and origin from the *in situ* studies accomplished by several multinational space missions, starting with Comet Halley in 1986 and going on to cover observations of Comet Hale-Bopp, the great summer-sky spectacle of 1997. Included along the way is the gravitational demolition of Comet Shoemaker-Levy 9 by Jupiter and the subsequent collision of the 20-odd icy fragments with the giant planet in 1994.

Although the missions to Halley had the obvious advantage of a close-up look and direct analysis of dust in the coma, traditional telescopic work has benefited enormously from the development of new and powerful instrumentation. Perhaps the most stunning advances have come from the application of radio astronomical techniques to comet studies. After a century of speculation on the identity of the parent molecules spawning all those complex photographic spectral lines and bands (from the daughter molecules), the parent molecules have finally come into view. Not only are water, carbon dioxide, carbon monoxide, methanol, isocyanacetylene and a host of new molecules detectable, but the distribution of the isotopes of hydrogen, carbon, oxygen and nitrogen that comprise them have become decipherable. In the isotopes lie clues to the ultimate origin of comet material.

Comet Science not only lays out the factual information we have about comets, but clearly describes the context in which we now regard these icy "remnants from the birth of the Solar System". Parallel advances, from theoretical and laboratory studies, in understanding the chemistry of the ice and dust in the interstellar medium that floats around and clumps up in regions separating the stars show that some cometary material originated before the Solar System began to form. Mineral grains in thin, dusty clouds in deep space acquire icy coatings that are chemically processed by ambient stellar ultraviolet light, marking the beginnings of the complex carbon chemistry that, transported in the comets that have hit Earth, seeded our planet, as well as uncounted others, with some of the raw materials of life.

Comets are routinely observed from space by the Hubble Space Telescope, which offers views unachievable from Earth, and much new information about the nature of water in comets was revealed by observations of



Spectacular apparitions: but comets may also reveal much about the origins of life.

Hale-Bopp with the European Space Agency's (ESA's) Infrared Space Observatory. More space missions to visit comets are in progress, such as Stardust, launched last year for a planned encounter with Comet Wild 2 in 2004. Others are in advanced preparation, including Rosetta, the ESA mission scheduled for launch in 2003 to meet periodic Comet P/Wirtanen in 2012–13, and the US space agency NASA's Deep Impact, scheduled for launch in 2004 and an encounter with Comet P/Tempel 1 in 2005.

Crovisier and Encrenaz describe the newly discovered Kuiper Belt and the more distant Oort Cloud, reservoirs respectively of the short-period and long-period comets — those with orbital periods of less or more than roughly 200 years. Dynamical studies are now telling us how comets are plucked from their distant orbits and drawn towards the inner Solar System, where the Sun's heat causes them to evaporate and glow in a brief but glorious passage through our field of view.

Comet Science is comprehensive yet brief, engaging yet authoritative, and fully accessible to a wide range of readers. It is well suited to student use and sufficiently detailed for astronomers wanting an overview of the state of the field, while remaining readable by the informed layperson. The translation from the original French is very smooth, and this edition is enhanced with colour pictures of comets and many technical diagrams.

The authors have captured the full panorama of the investigations of these primitive lumps of dirty ice, the messages they carry concerning the origins of the planets, and perhaps the origin of Earth's teeming veneer of organic matter, that includes the likes of us. ■

Dale P. Cruikshank is at the NASA Ames Research Center, Moffett Field, California 94035-1000, USA.

More on the cosmos

Hubble Space Telescope: New Views of the Universe

by Mark Voit

Harry N. Abrams, in association with the Smithsonian Institution & the Space Telescope Science Institute, \$19.95, £12.95 (pbk)

Unfolding Our Universe

by Iain Nicolson

Cambridge University Press, £24.95

Other Worlds: Images of the Cosmos from Earth and Space

by James

Trefil

National

Geographic Society, £22.50, \$35



Science in culture

Nature's microscopic art forms

Radiolarians and diatoms drawn by Ernst Haeckel.

Philip Ball

Design in nature, a source of inspiration both to the artist and the engineer, has also solicited the attention of philosophers in the past. Where do the stunning beauty and elegance of the microscopic exoskeletons of radiolaria and diatoms come from? These works of natural micro-architecture were revealed for the first time to an astonished scientific community in the late nineteenth century, when the British research vessel *HMS Challenger* collected sediments from the seabed. The wonders of *Challenger's* booty were painstakingly drawn and catalogued by the German biologist Ernst Haeckel.

Haeckel's legacy to art, science and design is explored in the *More Than Meets the Eye* exhibition currently at the Victoria and Albert Museum in London (see *Nature* 407, 20; 2000).

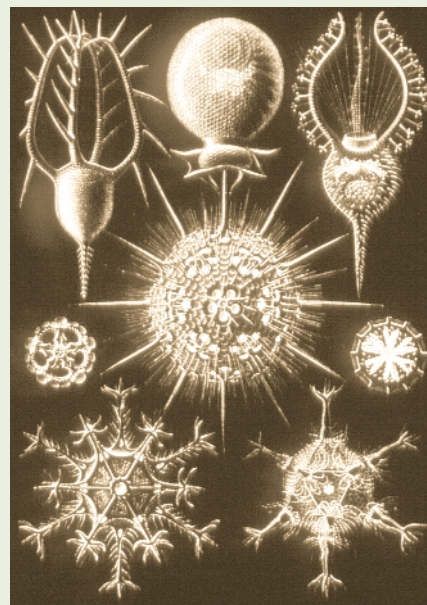
The exhibition features a display that looks at biomimetic elements of art and engineering, ranging from art nouveau to the geodesic domes of US architect Richard Buckminster Fuller and the 'soap-film', tent-like structures of German architect Frei Otto.

Buckminster Fuller's domes are echoed in the spherical exoskeleton of the radiolarian *Aulonia hexagona*, one of those lovingly drawn by Haeckel for his portfolio *Art Forms in Nature*, published between 1899 and 1904. Haeckel's illustration shows the occasional pentagons in the web of hexagonal cells, necessary (as eighteenth-century Swiss mathematician Leonhard Euler showed) to produce curvature and close the sphere.

These radiolaria were not the first microscopic 'art forms' to be identified in nature. Christian Ehrenberg observed delicately patterned coccolithophore shells in chalk in 1836 — but concluded that they were inorganic formations. By that time, studies by biologists, including Thomas Huxley, had convinced most others that coccoliths were biogenic.

Yet the range and inventiveness of designs in radiolarians and diatoms were quite unprecedented. Such discoveries inspired a strong interest in natural form in the early twentieth century, most famously expressed in D'Arcy Wentworth Thompson's *On Growth and Form* (1917). Yet while the Scottish naturalist was busy showing in meticulous detail that these 'watches' needed no maker beyond the laws of physics and mechanics, his compatriot James Bell Pettigrew at the University of St Andrews was at work constructing a kind of theistic argument from design in his *Designs in Nature* (1908). Pettigrew saw the imprint of continual intervention from the divine in the geometric forms of nature.

Both men had to come to terms with what Haeckel himself had said on the matter. In Haeckel, science, art and philosophy combined to produce a heady brew, strongly flavoured by the



Computer-enhanced examples of Haeckel's painstaking work.

German *Naturphilosophie* instigated by Johann Wolfgang von Goethe (whom Haeckel idolized). Haeckel is perhaps best regarded as a mystical atheist, sharing a Nietzschean belief that "Man creates God in his own image" and yet prepared to embrace the idea of a life-giving spirit that permeated matter even down to the atomic level.

This outlook is encapsulated in Haeckel's book *Crystal Souls* (1917), in which he presciently attempts to link the geometry of radiolaria with the discovery of liquid crystals by Friedrich Reinitzer in 1888. Otto Lehmann, the German physicist who helped the botanist Reinitzer make sense of his discovery, suggested in 1904 that liquid crystals were intermediate between living and non-living matter. We now know that liquid crystals can indeed organize themselves spontaneously into ordered aggregates that provide templates for the formation of inorganic materials regularly patterned on the microscopic scale, and that this mimics some of the processes of biomineralization that occur in radiolaria.

Haeckel's contributions came at a time when Darwin's evolutionary theory was provoking debate both about the origin of life and about political theory: Marx and Engels were grappling with its implications. Blended with the nationalism that was in the air in Germany and with Haeckel's political ambitions, this lends some of Haeckel's views — such as those on eugenics — the taint of racial élitism from which *Naturphilosophie* has never been entirely free.

Yet his gorgeous illustrations had a wide impact, reaching the eyes of the art nouveau artists and their counterparts, the *Jungenstil* group, in Germany. Nature's ornamentation blossomed throughout Europe. ■

Philip Ball is a consultant editor at *Nature*.

A victim of truth

Vitalism was an attempt to reconcile rationality with a sense of wonder.

Sunetra Gupta

Convictions, said Nietzsche, are more dangerous enemies of truth than lies. Among the catalogue of convictions that the human race has challenged and eventually relinquished is the fascinating notion that living organisms are distinct from nonliving entities by possessing a 'vital force'. The philosophy of 'vitalism' has its roots in the original distinction between organic and inorganic compounds, dating from around 1600, which was based on their reaction to heat. Although both types of substance changed form when heated, inorganic compounds could be recovered upon removing the heat source, whereas the organic compounds appeared to undergo a mysterious and irrevocable alteration.

The implication that the latter were imbued with a vital force gave birth to an idea that eventually came to occupy a very tricky position between materialism and idealism by endorsing the viewpoint that, although organic material might obey the same physical and chemical laws as inorganic material, life could not be governed by these laws alone.

Vitalism's singular place in history rests on its attempt to reconcile two opposing needs — the need for analytical reasoning and the need to celebrate the mystery of human experience. The life of the Swedish chemist Jons Jacob Berzelius (pictured on the right) traced the tensions between these concerns in dramatic detail.

We are accustomed to thinking of the defining event in Berzelius's life as the letter from his student that declared: "I must tell you that I can make urea without the use of kidneys, either man or dog." The year was 1828 and Friedrich Wohler, in setting out to synthesize ammonium cyanate, had obtained a white crystalline material which proved identical to urea. It was the first organic compound to be synthesized from inorganic starting materials, and the achievement knocked down one of the few

It seems that we are still trapped in a state of poetic ambivalence as to whether life is greater than the sum of its parts.

The life of Jons Jacob Berzelius traced the tensions between materialism and idealism in dramatic detail.

remaining tenets of vitalism — that although organic chemicals could be modified in the laboratory, they could only be produced through the agency of a vital force present in living plants and animals.

Berzelius apparently tried to downplay Wohler's discovery by exiling urea to a hinterland between organic and inorganic compounds. An alternative school of thought proposes that his lack of enthusiasm had more to do with the problems it posed for his own theory of inorganic compound formation. At any rate, the event struck many, including Wohler himself, as being more remarkable in demonstrating how a salt (ammonium cyanate) could reconstitute itself into an organic substance with the same empirical formula.

By deflecting the issue towards the structural implications, Berzelius was able to maintain a dignified silence on the question of vitalism, leaving us with room to speculate about what the discovery might really have meant to him personally. During his long encounter with chemistry he vacillated between stances that are clearly supportive of a mystical vitalist force and others that are more accommodating of an atheistic materialism, which he generally abhorred. It appears that much of his energies as a chemist were engaged in the honest negotiation of a compromise between these two poles.

Vitalism did not die with the synthesis of urea, but its boundaries were pushed back a little further. Vitalists now began to contend that it was an organism's functioning rather than its constituent substances that lay outside the boundaries of human comprehension. Berzelius, in his final analysis, acknowledged that the notion of a vital force as distinct from normal inorganic forces was invalid; instead, organisms were to be distinguished by a mysterious arrangement of "circumstances" dictated by the (divine) purpose of producing life. By the early twentieth century, the focus of vitalism had shifted to



MARY EVANS

another set of circumstances — namely, the development of an organism. Known as entelechy, the concept that a vital force accounts not only for the maintenance of life but also for its development was used by the vitalist Hans Driesch to explain the astonishing process of embryonic differentiation.

Are we where we are today in this process of gradual erosion? We now have a 'working draft' of the human genome and still the engineers of such a feat are anxious to emphasize "the imponderables of the human spirit". It seems that we are still — perhaps happily so — trapped in a state of poetic ambivalence towards the question of whether life is greater than the sum of its parts. Like Berzelius, we remain inclined to believe that the analysis of life does not detract from its ultimate mystery. ■

Sunetra Gupta is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Win a Nobel prize!

Wealth, chicks, guys, the secrets of the Universe — they can all be yours.

Vernor Vinge

Dear Johann — I was sorry to learn that you have been passed over for tenure. I hope you won't give the bums on your committee another chance to abuse you.

This was going to be an ordinary letter. Then I realized that you probably don't remember [me](#). We took the same section of Fong's Comparative Genomics class at Berkeley, but I quit the program and drifted into the arts (see my [SensationXXX](#) performances). Now I work in human resources. It's a perfect fit for my technical and people skills.

Johann, what I have to offer you is so extreme that I'm afraid your filters would trash my mail before you ever see it. That's why you are reading this as an advertisement in your personal copy of *Nature*: hopefully, this will show that we're serious.

In fact, writing this ad has been a lark. Yes, it's over the top ... but it's also the absolute truth. Working with us, you can win a Nobel prize, and that is just the beginning. So in just a few words, I have to convince you to take the next step.

I know you read outside your field, Johann. That's one reason why unimaginative drudges get tenure and you don't. Have you been following the news about MRI-with-transfection? The enabling mechanism is an HIV transfection of the subject's glial cells. The inserted genetic material expresses proteins which can be signalled by a ten-gauss modulation of the MRI's gradient magnets. Synched with the rf pulses, they promote the production of selected neurotransmitters. If it's done right, the experimenter can trigger from an alphabet of about 20 neurotransmitters — at a spatial resolution only twice as coarse as the MRI's imaging resolution.

The neuroscience guys have fallen in love with this. And right behind them are the psych people: with whole-brain MRI/t scans, researchers could induce almost any psychopathology. In public, that possibility is just ominous speculation. In secret, at least three research labs already have whole-brain MRI/t. We have such systems ourselves, and although we haven't abused them, they are more scary than the editorials. One of the most horrifying mind-sets is something we call 'specialist fugue state'. When applied to a researcher, it creates an idiot savant, without a life beyond short-range research goals.



This is not what I'm selling you, Johann! But beware. Several labs are recruiting specialists for just this nightmare. Maybe they're getting fully informed volunteers; more likely they are getting duped victims. Either way, the public will soon be seeing all sorts of research productivity that is secretly based on this modern form of slavery. Don't you get trapped by such a scam.

No, if you work for us, you'll be running the biotech show. Johann, you are brilliant and well trained and ... well, we've studied you pretty carefully. You can name your price. We have major financing from a small but wealthy nation state. If you buy in, you'll have resources that rival the CDC: a ten-petaflops computer with a storage area network that mirrors the largest dynamic proteomics sites. All this — and the support staffs — will be fully dedicated to your personal use.

So what's our secret? Well, we've improved the MRI/t trigger mechanism to respond on millisecond timescales. We can induce direct brain I/O with the look and feel of memory and thought. For 50 years, people have been predicting mind/machine symbiosis. Now we've actually done it, Johann! You'll want to talk to Wardner. He's our first success, a perfect fit for the technique, although his specialty is strategic planning. With our MRI/t technique, Wardner is like a god.

You know how your field is these days: more breakthroughs than ever before — but it's dull, dull. A modern cell-mechanics lab is like an old-time genomics site — a quietly humming data factory. The same thing has happened in the non-bio sciences. Some theoreticians think this is heaven, but take a look at the [2013-01-17 editorial](#) in *Nature*:

for every breakthrough, there are a thousand more hiding in the new databanks.

You can change that, Johann. Your mind will interact directly with our world-class automation. You'll solve protein dynamics problems as easily as ordinary people plan a day at the beach.

Your working conditions? They can be almost anything you want — except that you'll have to relocate. We've already built a large [villa](#) for you at our [Riviera research site](#). You'll have complete freedom of movement. The transfection is not reversible, but it's easy to 'safe' the neuroactives when you are not actually connected. And of course, being 'connected' doesn't involve any messy electrodes. You simply enter the study that we've built in your villa. It's quite spacious, considering it's inside a four-tesla MRI system. (And we must be very careful about magnetic materials; Wardner can tell you the usual bozo stories about high-velocity jewellery.)

Well, that's my pitch, Johann. Obviously, our company must be very secretive at this stage. But please, come out and visit us. No obligation, except to sign a nondisclosure agreement. We ask that you don't tip off your colleagues about this short visit, but we want you to be absolutely comfortable about it. You have family, a cousin I believe? Feel free to let her know where you are going.

I hope you can come, Johann.

Your friend,

Helen

Helen Peerless,

Director for Human Resources, [Mephisto Dynamics](#)

From 1972 to 2000, Vernor Vinge taught mathematics and computer science at San Diego State University. His most recent novel is A Deepness in the Sky (Millennium).

Making a potential difference

George E. Blomgren

To reduce our use of fossil fuels we will need cheap and safe batteries to run electric vehicles and store energy — magnesium batteries may be the answer.

On page 724 of this issue¹, Aurbach *et al.* describe a prototype for a rechargeable magnesium battery, which may eventually replace the standard batteries used in large-scale energy conversion. Such a system has been a long-sought goal in electrochemistry because it could provide a cheap and environmentally benign alternative to existing options, and may help to reduce the dependence of modern societies on fossil fuels.

Batteries are at the heart of the portable devices of the information age. In the past decade the lithium-ion battery has become the main source of energy for these devices, replacing older nickel–cadmium and nickel–metal-hydride varieties. The past year has also seen the introduction of lithium-ion polymer batteries — in portable telephones and handheld personal digital assistants, they are replacing nickel and standard lithium-ion batteries.

Despite these rapid developments in portable technology, better and cheaper versions have not yet replaced the standard large batteries used for 'heavy load' applications, such as stationary energy storage. The lack of suitable alternatives has compromised the progress of replacement technologies for hydrocarbon fuels. For example, growth in the use of solar power has revealed a need for better batteries to store energy for use after the Sun goes down. The development of electric vehicles has likewise been held up (Fig. 1). An ordinary lead–acid battery is made from cheap and abundant, but heavy and highly toxic, materials, whereas newer nickel batteries contain toxic cadmium or scarce rare-earth-hydride materials. Even the more expensive lithium systems are reactive when exposed to air. So they all have serious flaws.

The subject of Aurbach and colleagues' study¹ — magnesium — is a light and plentiful element found in brines and minerals worldwide. It is environmentally benign and safe to handle. For these reasons, many electrochemists have investigated magnesium as an anode material for batteries, and several companies have produced primary (not rechargeable) batteries, mainly for military applications. The problem in developing rechargeable batteries with magnesium anodes is finding the right electrolyte.

All batteries have an anode (negative electrode), a cathode (positive electrode) and an



Figure 1 No longer just a concept car? Better rechargeable batteries are needed to run electric vehicles, such as this Nissan Hypermini. Because existing battery systems can provide only enough energy for short trips, electric vehicles are often designed to be small or made from lightweight materials to reduce the load on the motor. So far, hybrid electric vehicles, which use an internal-combustion engine for longer distances, have made the most progress — but they still emit global-warming pollutants. Cheap, non-toxic and lightweight batteries, such as the magnesium system developed by Aurbach *et al.*¹, are essential to make electric power competitive with existing combustion engines.

ion-conducting electrolyte. The electrolyte forms a continuous phase between the two electrodes to allow ions to pass and complete the electric circuit. Electrons flow through the external circuit to do the electrical work. It is important that most of the bulk volume of each electrode is chemically active and in contact with the electrolyte. The problem with most electrolytes used with magnesium anodes is that they allow a surface film to grow on the anode, stopping the electrochemical reaction.

For a battery to be rechargeable it must be able to completely reverse the direction of the electrochemical reactions at each electrode, from charge to discharge and vice versa. During discharge a magnesium electrode gives up electrons to the external circuit, and the resulting magnesium ions pass into the electrolyte. During charge, the magnesium ions are plated back as metal onto the

negative electrode, which also takes up electrons. For the anode, these two processes are expressed succinctly by the electrochemical equation: $\text{Mg} \rightleftharpoons \text{Mg}^{2+} + 2\text{e}^-$, where the reaction goes in the forward direction during discharge and the backward direction during charge, and e^- is an electron. Similarly, the reactions at the cathode may be represented as $\text{Mg}_x\text{MX} + z\text{Mg}^{2+} \rightleftharpoons \text{Mg}_{x+z}\text{MX} + 2\text{e}^-$, where Mg_xMX represents a magnesium-containing compound. The challenge has been finding a suitable cathode material that can reversibly bind magnesium ions and remain stable.

Electrolyte solutions, such as Grignard reagents, have long been known to deposit magnesium reversibly². Grignard reagents have the formula RMgY , where R is an alkyl or aryl group and Y is chlorine or bromine in ether solution; they are commonly used for alkylation reactions in organic synthesis. But they are strong electron donors, so they reduce and destabilize the cathode. Part of the success of Aurbach and co-workers' study was finding an electrolyte solution that can reversibly deposit magnesium without destroying the cathode.

The authors used electrolytes based on magnesium–organohalo aluminate salts (actually two types: $\text{Mg}(\text{AlCl}_3\text{R})_2$ and $\text{Mg}(\text{AlCl}_2\text{RR}')_2$, where R and R' are alkyl groups) with a solvent of cyclic ether (tetrahydrofuran) or a polyether of the glyme family such as diglyme. Ether solvents like these have been widely studied in lithium systems and are known to be stable. But the electrolyte salts are new to battery research. Aurbach *et al.* chose to use the aluminate form to improve cathode stability and magnesium deposition. This type of electrolyte will form a battery with a voltage as high as 2.5 V, much better than aqueous electrolytes (which are thermodynamically stable to just over 1 V) and high enough for commercial interest. The voltage range is not as good as with non-aqueous electrolytes used in lithium batteries, but those electrolytes would not reversibly deposit magnesium.

The second part of the discovery was to find a suitable material for the cathode. The authors found that $\text{Mg}_7\text{Mo}_3\text{S}_4$ would reversibly accept and release magnesium ions with a cell voltage of about 1.5 V. Nickel battery cells operate at 1.2 V, whereas lead–acid cells operate at about 2 V, so the cell is competitive with these batteries. Early

CORBIS

results from tests of this material in electrolyte cells have shown an acceptable lifetime, with a competitive energy per unit weight at reasonable rates of charge and discharge. Based on material costs, it is anticipated that the system might fall between aqueous batteries and lithium-ion batteries in terms of price.

As with any new technology, many tests must be done to fully characterize the system and to optimize the battery's behaviour, and these will no doubt be carried out by battery manufacturers with an interest in the system.

In the end, of course, it must eventually prove itself in the marketplace. But our need to find viable alternatives to hydrocarbon fuels means there is every incentive to make full and fair evaluations of technical advances such as this one.

George E. Blomgren is at Blomgren Consulting Services Ltd, 1554 Clarence Avenue, Lakewood, Ohio 44107, USA.

e-mail: geblomgren@prodigy.net

1. Aurbach, D. et al. *Nature* **407**, 724–727 (2000).

2. Connor, J., Reid, W. & Wood, G. J. *Electrochem. Soc.* **104**, 38–41 (1957).

Computational neuroscience

Building blocks of movement

Zoubin Ghahramani

When we walk, turn the pages of a book, or catch a ball, our brain is solving a 'control problem' — that of coordinating the activities of all the muscles required to achieve these voluntary movements. Roboticians have long appreciated the computational challenge of controlling such movements. For even a simplified description of the human arm, several pages of equations are required to express the forces that muscles need to exert to accelerate the arm in a given direction. And these equations would change when the arm picks up an object or moves in a different medium,

such as water, or as muscles tire. The challenge for neuroscientists has been to discover how these equations are represented in the brain, and how these representations adapt to suit new environments or tasks. On page 742 of this issue, Thoroughman and Shadmehr¹ provide some answers.

They do this by expanding on an idea that has captured the imagination of physiologists, modellers and psychologists alike. This idea is that the central nervous system creates seemingly complex behaviours by combining a relatively small number of simpler 'motor primitives'. Like a child putting

together simple building blocks to create elaborate structures, the brain — so the theory goes — puts together these motor primitives to create new behaviours.

What might constitute these 'building blocks' of movement? Many (not mutually exclusive) suggestions have been made, such as synergistic muscle contractions², elementary stable postures³ and simple movement strokes⁴. But the problem for experimentalists has been to establish the existence of these primitives, to obtain a 'picture' of what they look like, and to understand their role in everyday behaviours. Thoroughman and Shadmehr¹ now provide a mathematical description of the motor primitives that the central nervous system uses to learn to control the arm.

To do this, the authors make use of a well-studied⁵ task in which subjects make reaching movements while holding on to a handle attached to the end of a lightweight robotic arm (Fig. 1a). When the robotic arm is turned off, this is a very easy and natural task. When the robotic arm is turned on, however, it generates forces that simulate a very strange kind of environment, not previously experienced by the subjects. In this environment, a reaching movement in any direction by the subject causes forces perpendicular to the direction of movement and proportional to the velocity of movement, transforming previously straight movements into roughly spiral-shaped ones. It is well known that subjects eventually adapt to this type of force field, but Thoroughman and Shadmehr study this adaptation in detail. Their ingenious analysis reveals remarkable structure in the pattern of adaptation, giving us a glimpse of motor primitives in action.

The authors reason that if a reaching movement occurring within a force field causes the brain to learn, and if this learning is represented by changes to motor primitives, then changes will be seen in subsequent reaching movements — not only in the target direction but also in other directions. So, by working out how learning 'spills over', or generalizes, between movement directions, the mathematical shape of the motor primitives can be uncovered.

Indeed, Thoroughman and Shadmehr find that individual movements do have an effect on subsequent movements in other directions. This generalization decays with the angular distance between movement directions: learning how to move the arm in one direction results in partial learning of how to move the arm in nearby directions, but unlearning of movements in the opposite direction. The pattern of generalization suggests that the motor primitives map the desired velocity of the reaching hand to the force required to move the hand at this velocity. The pattern also suggests that mathematically the primitives have a gaussian shape (Fig. 1b).

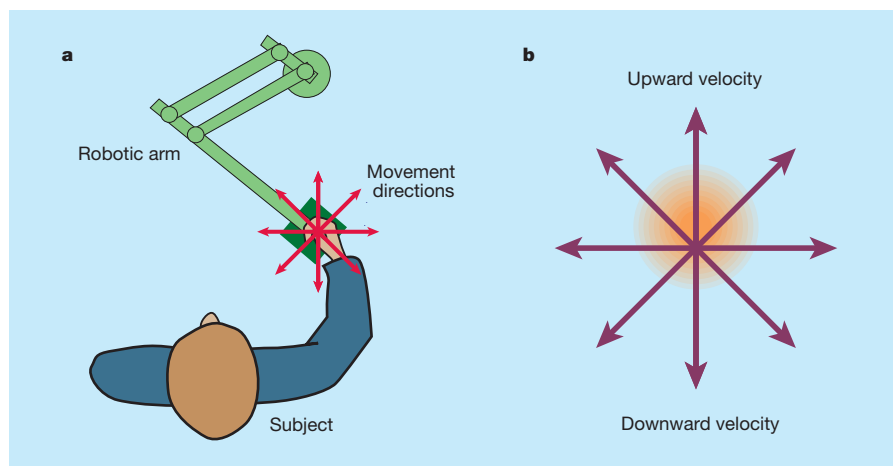


Figure 1 Producing complex movements from simple building blocks. Thoroughman and Shadmehr¹ investigated 'motor primitives', thought to represent the neural building blocks of complicated movements. a, The experiment. Human subjects held on to a robotic arm and moved their hand towards targets in eight different directions. The robotic arm exerted forces on the hand that depended on the velocity of movement of the hand. At first these forces affected the reaching movements by the subjects, but eventually the subjects adapted to the forces, allowing them to reach in the correct direction. b, The model. The pattern by which a reaching movement in the force field in one direction affected subsequent movements in other directions suggested a model in which the dynamics of movement are learned by changing and combining motor primitives that encode the force required to move the arm at a particular velocity. The primitives overlap (only one is shown, in orange), and have a broad gaussian tuning to desired hand velocity. The primitive shown is tuned to produce the force needed for movement in the upward direction, but spills over into other directions in a manner that matches the experimental results¹.

The authors then devised an adaptive, computational model composed of such primitives, and compared the behaviour of humans and the model in several new variations of the task. The model correctly predicted the S-shaped trajectories of human reaching movements; that these S-shaped trajectories would become straight if the proportion of 'catch' trials (when the robot was surreptitiously turned off) were decreased; and how subjects would adapt incorrectly to force fields that changed direction too rapidly.

These results are interesting for three reasons. First, it has been suggested⁶ that neither the inputs ('proprioceptive' neuronal signals) nor the outputs (muscle activations) of the human movement-control system show a simple relationship to external forces or the speed of movement of a reaching hand. So it is remarkable that Thoroughman and Shadmehr's model — in which motor primitives map desired hand velocity to force — can account for the details of the time course and the end product of the human subjects' adaptation. This type of mapping is exactly what makes control through motor primitives computationally attractive. The neural 'controller', presumably in the brain and spinal cord, need not concern itself with how hand velocities are computed or how forces are produced; its task is simply to learn how to map between the two.

All the same, we should not jump to conclusions from these appealing results. Motor primitives that appear to encode extrinsic variables, such as force and hand velocity, are somewhat divorced from the nitty-gritty of the muscle-skeletal system. Such primitives might turn out to be encoded instead in terms of more fundamental variables, such as muscle activations⁷.

Second, we already knew that the computations required to adapt to new, dynamic environments need to be implemented somewhere in the central nervous system. Thoroughman and Shadmehr's results implicate one brain region — the cerebellum — as the area responsible. In particular, their finding that the motor primitives are velocity-tuned fits in with the encoding of velocity seen in Purkinje cells in the cerebellum. Other evidence also points to a key role for the cerebellum in motor learning. But of course things are not clear cut — previous studies using a different generalization paradigm⁸ and primate neurophysiology⁹ indicated that the brain region responsible for learning new movement dynamics could be the primary motor cortex. It is likely that changes in both cerebellum and primary motor cortex are involved in learning this particular task. Besides which, there is little reason to think that a computational element such as a motor primitive should map simply to brain structure.

Finally, these results are interesting from

a robotics perspective. Most artificial learning systems sluggishly accumulate small changes. But in humans, a single movement within the force field can have significant effects on subsequent movements — this biological movement-control system remains stable while adapting rapidly. By studying the building blocks of movement we not only learn more about the neurobiology of movement in humans. Some day, we may want to put those building blocks into robots, too.

Zoubin Ghahramani is in the Gatsby Computational Neuroscience Unit, University College London, London WC1N 3AR, UK.

Conservation biology

Seeds of doubt

Peter D. Moore

What is a weed to a farmer may well be a cherished wild flower to someone else, or an essential ecological component of a local flora. Hence the attempts, in response to diminishing botanical diversity, to sustain some wild plants by sowing their seeds in areas where they are under threat. But as Keller and colleagues show in *Journal of Applied Ecology*¹, this kind of conservation work can have adverse effects by diluting and modifying the local genetic resources of residual weed populations.

Agricultural weed control methods have generally met with considerable success, with the result that crops are more productive and harvests are less contaminated with unwanted seeds. But the side effects of weed decline include changed rural landscapes, falling numbers of invertebrates and even decreases in the populations of some birds. Weed seed mixtures are now commercially available (marketed, of course, as wild flowers), and an understandable response of environmentalists has been the deliberate dispersal of such seeds, especially along road margins in Europe and North America.

On the face of it, the re-establishment of more biodiverse agricultural landscapes is a creditable and straightforward aim. Simply allowing nature to take its course, by stopping herbicide treatment or simple disturbance of soil, or both, might lead to weed resurgence, but only of those that are well represented in the soil seedbank. Some weeds have seeds with a limited capacity to survive in soil, and these species in particular are candidates for supplementation with seeds from elsewhere. One disadvantage with this approach is that introduced plants, even those of the same species as local plants, might be poorly adapted to local conditions, including both the physical conditions of the habitat and the indigenous grazers, polli-

e-mail: zoubin@gatsby.ucl.ac.uk

1. Thoroughman, K. A. & Shadmehr, R. *Nature* **407**, 742–747 (2000).
2. Bernstein, N. *The Co-ordination and Regulation of Movements* (Pergamon, Oxford, 1967).
3. Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. *Science* **253**, 287–291 (1991).
4. Morasso, P. & Mussa-Ivaldi, F. A. *Biol. Cybern.* **45**, 131–142 (1982).
5. Shadmehr, R. & Mussa-Ivaldi, F. A. *J. Neurosci.* **14**, 3208–3224 (1994).
6. Loeb, G. E. in *Progress in Brain Research* vol. 123 (ed. Binder, M. D.) 405–409 (Elsevier, Cambridge, 1999).
7. Todorov, E. *Nature Neurosci.* **3**, 391–398 (2000).
8. Goodbody, S. J. & Wolpert, D. M. *J. Neurophysiol.* **79**, 1825–1838 (1998).
9. Gandolfo, F., Li, C.-S. R., Benda, B. J., Padoa-Schioppa, C. & Bizzi, E. *Proc. Natl Acad. Sci. USA* **97**, 2259–2263 (2000).

nators and pathogens. Can maladaptation of this sort be transferred to the residual local populations of plants?

To test this possibility, Keller *et al.*¹ studied three weed species, the common poppy (*Papaver rhoeas*), corncockle (*Agrostemma githago*; Fig. 1) and white campion (*Silene alba*). Their seeds are available commercially from several countries, including Switzerland, Germany, Britain, Hungary and the United States, and all are derived from local wild populations. The authors conducted a range of crossing experiments (always using Swiss material as mother plants) and measured such features as biomass yield, sur-



Figure 1 The corncockle. This weed (or flower, according to your point of view) is now rare in Britain but is still relatively common in southern and eastern Europe.

HOLT/BOB GIBBONS

vivorship and seed mass over two generations.

Biomass was almost always greater in the first-generation offspring of crosses involving foreign material, but declined in the second generation. Mortality was higher in both generations. Seed weight, like biomass, was generally higher in the first generation and lower in the second generation of foreign crosses. So it seems that, in the long term, the introduction of foreign genes into these weed populations is likely to make them less fit — that is, less able to survive and reproduce effectively. Relative reductions in fitness were estimated to be between 8% and 23%. The overall message is that the introduction of genes from distant populations is likely to do lasting harm to the native weed flora of an area.

These findings have broader implications. The introduction of fresh breeding stock into fragmented and isolated populations is often seen as a way of increasing genetic diversity, and has been attempted in organisms ranging from butterflies to birds². The merits of such a policy can vary with the geographical distance from which introduced material is obtained and with the spatial variability in the genetic constitution of the species concerned. Geographical distance might not always correspond to genetic distance: it will also vary with the type of organism involved.

The effects of genetic distance have been tested on a Californian shrub, *Lotus scoparius*, the deerweed³, using degrees of enzyme variation as the genetic measure. Deerweed is a variable shrub of the coastal sage community of the west coast of North America, being found in both arid and well-watered habitats. In an experiment involving 12 populations, there was only weak correlation between geographical and genetic distances. But in transplant experiments the cumulative fitness of plants showed an inverse relationship to genetic distance. The closer the relationship is, the more likely it is that the individual will survive well in similar situations. So the introduction of this shrub should be determined by an analysis of genetic or ecological similarity (or both): geographical proximity of a seed source will not necessarily provide the most appropriate material for translocation.

A further complication with flowering plants arises from their different breeding strategies. The corncockle, for example, often self-pollinates, so its genetic constitution is likely to be patchy on a local scale. Deerweed is pollinated by insects, which might also restrict the distances over which outbreeding is possible.

How does this relate to wind-pollinated, outbreeding plants such as birch, poplar, alder and elm? A modelling exercise, based on data concerning the timing of flowering and its response to climate⁴, has shown that there is little evidence of local genetic

variation in these and other tree species across Europe. The production of large amounts of pollen and the potential for distant dispersal in these species ensures rapid gene flow and little opportunity for local isolation and adaptation. In such outbreeding species, one could argue, the problems of selecting appropriate stock for sowing or transplanting in new locations are less serious.

The history of plant and animal introductions is littered with catastrophes⁵. These reports concerning the implications of mov-

ing genes between populations suggests that caution is needed here also.

Peter D. Moore is in the Division of Life Sciences, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 8WA, UK.

e-mail: peter.moore@kcl.ac.uk

1. Keller, M., Kollmann, J. & Edwards, P. J. *J. Appl. Ecol.* **37**, 647–659 (2000).
2. Evans, E. I. *et al. Br. Birds* **90**, 123–138 (1997).
3. Montalvo, A. M. & Ellstrand, N. C. *Conserv. Biol.* **14**, 1034–1045 (2000).
4. Chiune, I., Belmonte, J. & Mignot, A. *J. Ecol.* **88**, 561–570 (2000).
5. Macdonald, I. A. in *Biodiversity and Global Change* (eds Solbrig, O. T., van Emden, H. M. & van Oordt, P. G. W. J.) 199–209 (CABI, Wallingford, Oxon, 1994).

Oceanography

Stirring times in the Southern Ocean

Sallie W. Chisholm

Almost half of the photosynthesis on Earth is carried out by phytoplankton in the sea. So these tiny cells play a huge part in the global carbon cycle, and in regulating climate by controlling the amount of

the greenhouse gas CO₂ in the atmosphere. Phytoplankton are the engine of the 'biological pump' (Fig. 1) that helps maintain a steep gradient of CO₂ between the atmosphere and deep ocean. It has been suggested that we

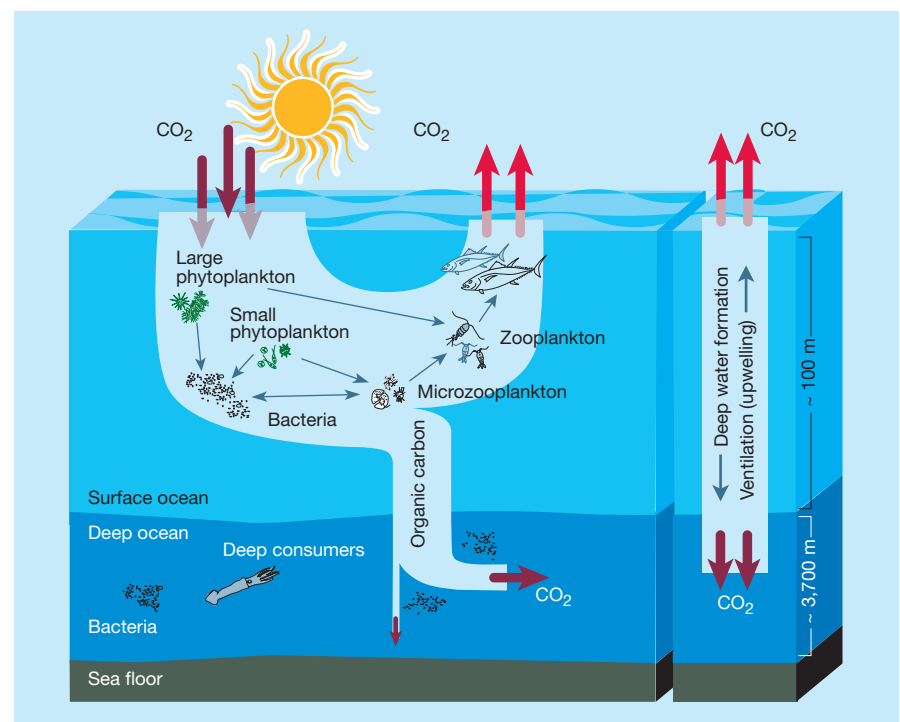


Figure 1 The 'biological pump' is a collective property of a complex phytoplankton-based food web. Together with the 'solubility pump' (right), which is driven by chemical and physical processes, it maintains a sharp gradient of CO₂ between the atmosphere and the deep oceans where 38×10^{18} g of carbon is stored. Using sunlight for energy and dissolved inorganic nutrients, phytoplankton convert CO₂ to organic carbon, which forms the base of the marine food web. As the carbon passes through consumers in surface waters, most of it is converted back to CO₂ and released to the atmosphere. But some finds its way to the deep ocean where it is remineralized back to CO₂ by bacteria. The net result is transport of CO₂ from the atmosphere to the deep ocean, where it stays, on average, for roughly 1,000 years. The food web's structure and the relative abundance of species influences how much CO₂ will be pumped to the deep ocean. This structure is dictated largely by the availability of inorganic nutrients such as nitrogen, phosphorus, silicon and iron. Iron is the main limiting nutrient in the Southern Ocean, which is why the SOIREE experiment^{1–3} was conducted there. (Figure modified from a graphic by Z. Johnson.)



100 YEARS AGO

The New York correspondent of the *Lancet* states that the Chicago Board of Education has established a department called "Child-study and Pedagogic Investigation". The examination is undertaken for the purpose of determining the mental and physical status of the school-children. Examinations were at first limited to the determination in each pupil of the following points: Height, height sitting, weight, ergograph work, strength of grip right and left, hearing right and left, and acuity of vision. In addition to this, obvious developmental defects have been noted. The number of children examined down to the present time is 5636. The conclusions thus far reached are that there is a physical basis of precocity, that dull children are lighter and precocious children heavier than the average child, and that mediocrity of mind is associated with mediocrity of physique... This is the first instance of a municipal board in America appropriating money for research work, and its effect may be far-reaching.

From *Nature* 11 October 1900.

50 YEARS AGO

There has probably never been a time in the history of the oil industry when 'petrol', to use the colloquial term, has engaged so forcibly public attention in Great Britain and elsewhere. Once rationing is imposed on any commodity which has been taken for granted, as up to 1939 petrol certainly was and likewise food (water has always been), there is solid basis for the ordinarily complacent citizen to know the reason why and to bestir himself to the facts. Even to keep a ration going for vital needs in time of emergency and stress is a formidable task; had it not been for amazing advances in technique of producing motor fuels by processes developed during the Second World War and afterwards, one wonders whether, indeed, economic factors aside, the petrol ration would not be with us in Britain and in Europe generally for years to come. The explanation of these advances is, of course, highly technical; it rests on the extraordinary rate of evolution in refinery procedure during the past decade and on conversion of petroleum in the more or less raw state into motor fuels by specialized thermal and catalytic processes. Dr. A. N. Sachanen devotes his large book to this complex subject and to the technologist. From *Nature* 14 October 1950.

might increase the efficiency of this pump — thereby drawing more CO₂ out of the atmosphere — by artificially supplying nutrients to the surface oceans. This suggestion is highly contentious. Papers by Boyd *et al.*¹, Abraham *et al.*² and Watson *et al.*³ (pages 695, 727 and 730 of this issue) will add fuel to the debate about the desirability of such 'geoengineering' solutions to Earth's ills. The papers describe the results of a fertilization experiment in the Southern Ocean, around Antarctica, and its scientific implications for interpreting past climate change.

Phytoplankton biomass in the global oceans can usually be correlated with the supply of nitrogen, phosphorus and silicon upwelling from the deep sea to the sunlit surface waters where photosynthesis takes place. In the equatorial Pacific and the Southern Ocean, however, these supplies far exceed demand, indicating that some other factor is limiting phytoplankton growth. This mystery plagued oceanographers for decades until the late John Martin showed that the limiting factor is iron, a trace element that reaches the oceans in atmospheric

dust⁴. Analyses of ancient ice cores show that dust deposition varied significantly during glacial and interglacial periods, and is anti-correlated with concentrations of atmospheric CO₂ over the past 400,000 years. So Martin reasoned that airborne iron supply could in part regulate climate by increasing the efficiency of the biological pump. He further suggested that very small amounts of iron distributed in today's Southern Ocean could draw a significant amount of CO₂ from the atmosphere. This is partly because the Southern Ocean is huge and, apart from iron, holds vast quantities of unused nutrients, and partly because its surface waters tend to sink, delivering carbon to the deep ocean.

With this as a backdrop, Boyd *et al.*¹ launched the SOIREE expedition (for Southern Ocean iron release experiment) in February 1999. They distributed 8,663 kg of an iron compound over a patch of ocean 8 km in diameter, some 2,000 km south-south-west of Hobart, Tasmania. The experiment was a dramatic success. Physiological indicators of iron stress decreased in the

Box 1 Commons concern

Whole-ecosystem experiments have revolutionized ecology⁹. With three successful ocean-fertilization experiments^{1,5,10}, oceanographers have joined that revolution. The discovery that iron limits phytoplankton growth in the equatorial Pacific and Southern Ocean has made it possible to stimulate the productivity of hundreds of square kilometres of ocean with a few barrels of fertilizer. By contrast, it would take roughly 3,000 times as much nitrogen and phosphorus to fertilize the North Atlantic, where these elements are limiting.

This is a powerful new tool for oceanographers, because by transient perturbation of the ecosystem from its quasi-equilibrium state we can glimpse the mechanisms that keep it there. These mechanisms, which consist of tightly coupled production and consumption processes in a complex food web (Fig. 1), hold one of the keys to understanding the connection between the oceans and the climate.

With powerful tools comes the responsibility of deciding

how to use them. Small-scale scientific experiments are one thing. But the correlations between iron availability, marine productivity and climate change have led to the prospect of using ocean fertilization to manipulate climate.

Coupled with the post-Kyoto possibility of a global market in carbon emissions trading, including the issuing of 'carbon credits'¹¹, this idea has spawned a budding industry^{12,13}. Patents on ocean fertilization have been issued to entrepreneurs and large corporations¹⁴, and there are plans for a 8,000-km² 'demonstration experiment'¹⁵. With a few exceptions¹⁶, little attention is being paid to the risks of large-scale fertilization. But we will be forced to make decisions about this technology, and quickly, whether we are ready or not.

In deciding on future uses of the ocean Commons, we have an opportunity to learn from past mistakes. The oceans are a complex adaptive system, so it is impossible to predict the long-term

consequences of commercial ocean fertilization. How then can we decide whether to proceed? We need to "reach for lessons"¹⁷.

The Earth system consists of elements distributed between the land, air and oceans by biological and geological processes over millions of years. Many environmental problems stem from our moving elements between these compartments at unprecedented rates that the system cannot accommodate. Importing massive quantities of nitrogen from the atmosphere to the land through fertilizer production, for example, has increased the production of the greenhouse gas nitrous oxide, and destroyed the ecology of coastal waters. Burning fossil fuels has moved massive amounts of carbon from the land to the atmosphere, and threatens to warm the globe. The direct and large-scale fertilization of the sea would undoubtedly have similar unintended consequences. The lesson is simple. In the long run, ocean fertilization is not sustainable. So why start? **S. W. C.**

iron-enriched patch within two days. Primary productivity, phytoplankton carbon and chlorophyll levels increased slowly but steadily over the 13 days that the patch was monitored, drawing nutrients and CO₂ from the surface waters. The result was a tripling of phytoplankton chlorophyll by the end of the experiment. These results are qualitatively much like those from a similar experiment⁵, conducted in the equatorial Pacific in 1995, but the development of the phytoplankton bloom was much slower and its magnitude smaller. The increase in phytoplankton biomass resulted in large part from a shift in the phytoplankton community from small-to large-celled species, primarily diatoms, which can outgrow their predators when suddenly released from nutrient limitation.

What was the long-term fate of the patch? From satellite pictures, Abraham *et al.*² demonstrate that a month after the end of the experiment the patch still had chlorophyll levels that were three times background concentrations, and had been transformed into a ribbon 150 km long and 4 km wide by stirring and diffusion. It stayed intact for at least two more weeks, having accumulated an estimated 600 to 3,000 tonnes of algal carbon. As the authors point out, however, there is no evidence that any of this carbon was exported to the deep ocean, as would be required for drawdown of atmospheric CO₂. In fact, over the period of the SOIREE measurements, the carbon export rates from the fertilized patch were actually lower than those of the surrounding waters, in part because diatom cells get lighter when released from iron starvation, and settle more slowly.

So SOIREE provides no evidence that iron fertilization increases carbon export from the surface to deep ocean. How then do these results relate to Martin's hypothesis that iron supplies to the Southern Ocean drive glacial-interglacial transitions? Watson *et al.*³ make this link using a model of the ocean-atmosphere carbon cycle to examine the effects of fluctuating iron supplies on atmospheric CO₂ concentrations over the past 400,000 years. They constructed the model using parameters drawn from SOIREE, 'forced' it with atmospheric iron fluxes derived from ice-core data, and compared the simulated atmospheric CO₂ concentrations with those measured in the ancient ice. The model reproduced the timing of glacial-interglacial fluctuations in atmospheric CO₂ remarkably well, challenging claims⁶ that the lag between the iron and CO₂ fluctuations in the ice core is too long for the relationship to be linked directly through Southern Ocean productivity.

One conclusion drawn by Watson *et al.* from their analysis is that "modest sequestration of atmospheric CO₂ by artificial additions of iron to the Southern Ocean is in principle possible", supporting Martin's

suggestion and the results of models designed to examine this possibility⁷. Although seductive in its simplicity, in practice this idea would threaten the ocean ecosystem (Box 1). Artificial fertilization with iron would probably have many unintended side effects, such as deoxygenating the deep ocean⁷ and generating greenhouse gases that are more potent than CO₂ (ref. 8). We know that it would change the structure of the marine food web.

Global biogeochemical cycles of elements have been shaped by the forces of evolution over geological timescales, and are tightly coupled. The injection of a single element (iron) into this system over a period of years to decades is not the same as the natural delivery of atmospheric dust to the oceans over thousands of years. Moreover, we cannot expect to change the flux of carbon from the atmosphere to the oceans — a single 'arrow' in this complex, self-organized system — without changing other features of the system in undesirable ways. It is likely that we would not recognize these changes until it was too late to reverse them. ■

DNA repair

Guarding against mutation

Richard D. Kolodner

Accurate duplication of a cell's DNA is essential for the viability and normal growth of that cell. But DNA replication occurs frequently and is complicated, so errors occur. This often results in incorrect pairing of the chemical bases that make up the rungs of a DNA ladder. If these mistakes are not fixed, harmful mutations can accumulate, so the error-correcting — 'mismatch-repair' — proteins are crucial. Defects can also occur in the genes encoding these proteins, resulting in high mutation rates and other damaging effects. For example, inherited defects in some human mismatch-repair genes cause a common syndrome characterized by a predisposition to cancer, and some sporadic cancers are due to defects in mismatch-repair genes that occur during normal cell growth¹. Key proteins in the mismatch-repair process are those that actually recognize the mispairs^{2,3}. The structural basis of this recognition process is revealed — for two bacterial proteins — by Obmolova *et al.*⁴ and Lamers *et al.*⁵ on pages 703 and 711 of this issue.

There are several mispair-recognizing proteins. In bacteria, these include the MutS protein — that studied by Obmolova *et al.*⁴ and Lamers *et al.*⁵. Here, two identical MutS proteins interact to form a functional 'homodimeric' complex. In eukaryotic organisms, such as humans, two different MutS-related proteins (such as MSH2 and

Sallie W. Chisholm is in the Department of Civil and Environmental Engineering, and the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: chisholm@mit.edu

1. Boyd, P. W. *et al.* *Nature* **407**, 695–702 (2000).
2. Abraham, E. R. *et al.* *Nature* **407**, 727–730 (2000).
3. Watson, A. J., Bakker, D. C. E., Ridgwell, A. J., Boyd, P. W. & Law, C. S. *Nature* **407**, 730–733 (2000).
4. Martin, J. H. in *Primary Production and Biogeochemical Cycles in the Sea* (eds Falkowski, P. & Woodhead, A. D.) 123–138 (Plenum, New York, 1992).
5. Coale, K. *et al.* *Nature* **383**, 495–501 (1996).
6. Broecker, W. S. & Henderson, G. M. *Paleoceanography* **13**, 352–364 (1998).
7. Sarmiento, J. & Orr, J. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
8. Fuhrman, J. A. & Capone, D. G. *Limnol. Oceanogr.* **36**, 1951–1959 (1991).
9. Carpenter, S. R. *et al.* *Science* **269**, 324–327 (1995).
10. Martin, J. A. *et al.* *Nature* **371**, 123–148 (1994).
11. Ney, R. A. & Schoor, J. L. *Env. Sci. Technol.* **34**, 177A–182A (2000).
12. Jones, L. S. F. & Otaegui, D. *Energy Conv. Mgmt* **38**, S367–S372 (1997).
13. Markels, M. *Regulation* **21**, 9–10 (1998).
14. www.uspto.gov Patent nos 6056919, 5992089, 5965117
15. Rawls, R. *Chem. Eng. News* 18 September, 66–69 (2000).
16. <http://esd.lbl.gov/DOCS/index2.html>
17. Goldenfeld, N. & Kadanoff, L. *Science* **284**, 87–89 (1999).

MSH6, or MSH2 and MSH3) interact to make up a 'heterodimeric' complex^{2,3}.

Proteins of the MutS family are remarkable sensors of DNA damage. The eukaryotic MSH2–MSH6 complex can detect several types of errors in DNA, with different consequences. For example, mispaired bases that arise as a result of DNA-replication errors are recognized by this complex, and repaired by mismatch repair^{2,3}. Another type of mispaired base — that resulting in reproductive and other dividing cells when similar chunks of chromosomes are swapped around by a process called recombination — is similarly fixed by mismatch repair^{2,3}. By contrast, when these recombining chromosome segments are too dissimilar, a different outcome ensues: the prevention of recombination^{2,3,6}.

Finally, eukaryotic MutS proteins can recognize chemical damage in DNA, including that caused by some drugs used for chemotherapy. This can activate cell-death pathways rather than DNA repair. Defects in this process result in cellular resistance to these drugs, and the resistance of cancer to chemotherapy^{7,8}. So, if we can unravel how the MutS proteins distinguish between so many types of problematic DNA structure, and communicate specifically with so many downstream pathways, we will not only gain greater insight into a fundamental biological process, but may also learn more about

stumbling blocks to the effectiveness of chemotherapy.

Several different models for the basic process of mismatch repair have been proposed, all of which conceivably include a 'signalling' element. In bacteria, MutS binds to a mispair and then interacts with MutL, signalling the activation of at least two other proteins². One of these proteins makes a break in the DNA strand containing the incorrect base. The other unwinds this segment of the DNA strand so it can be destroyed. The resulting gap in the strand will then be repaired. A similar process could be at work in eukaryotes, although it has also been suggested that the eukaryotic MutS complexes function as mobile signalling molecules⁹. These signalling cascades appear to rely on protein-protein interactions and conformational changes in proteins, which is why it is so important — in terms of understanding mismatch repair — that the structure of the first protein in the process has now been unveiled^{4,5}.

Obmolova *et al.*⁴ have been studying the structure of the MutS dimer from the bacterium *Thermus aquaticus*, bound to an unpaired base in a DNA strand. Lamers *et al.*⁵, meanwhile, looked at the *Escherichia coli* MutS dimer, bound to a mismatched base pair. Remarkably, even though the two MutS proteins that make up each dimer are identical, the dimer assumes an asymmetric conformation on binding the mispair, forming a ring around the DNA. Only one of the subunits contacts the mismatch, although both subunits have specific contacts with DNA. Recognition of the mispair by the dimer leads to a bending of the DNA.

Each MutS protein has a region that binds and hydrolyses ATP, producing ADP. It seems^{4,5} that these regions also differ between the two MutS proteins. The ATP-binding domains interact with each other and with the DNA-binding regions of both proteins, allowing the effects of ATP/ADP binding and DNA binding to be transferred across the protein. These details provide a structural basis for the biochemical observation that the binding and hydrolysis of ATP results in conformational changes in the MutS protein and in the modulation of protein-DNA interactions.

An intriguing question is how the MutS proteins bind specifically to DNA containing mismatched bases. These proteins have only about a 20-fold higher affinity for mispaired bases relative to normal base pairs, and mispaired bases are not all that common — they occur probably only once in every one million to ten million base pairs¹⁰. The new structures suggest one possible answer: that only binding to a mispair induces the conformational changes seen in both MutS and the bound DNA. If these conformational changes are required for the MutS proteins to interact with other proteins during mis-

match repair, then these changes and subsequent protein-protein interactions may effectively increase mispair specificity.

The eukaryotic mismatch-recognizing dimers — each of which consists of two different MutS-related proteins — are also asymmetric, in that the two subunits are related but different. The structures of the eukaryotic dimers have not yet been resolved. But the obvious parallels between the eukaryotic and bacterial complexes offer the opportunity for comparative structure-based biochemical studies, and may also be useful in the clinic.

We know that some heritable mutations in the human MSH2 gene predispose patients to a cancer known as hereditary non-polyposis colorectal carcinoma, or HNPCC. But a vexing problem is how to interpret the variety of amino-acid-altering mutations that have been found in the MSH2 gene in people suspected of having this syndrome. The available biochemical, clinical and population data are often insufficient to allow researchers to state with confidence that a particular sequence variation indeed causes disease, or is instead simply a natural variation in the human population¹¹. In general, it has not yet been possible to test directly the effect of a particular amino-acid-changing mutation on the function of the MSH2 protein. So, not surprisingly, clinicians are reluctant to make decisions about treatment on the basis of finding a

given variation in MSH2 in either cancer patients or healthy individuals suspected of having HNPCC.

But the two new crystal structures^{4,5} will provide a valuable starting point for molecular models of the effects of mutations in the MSH2 gene. The results of the simulations will guide studies of MSH2 function, and help researchers to work out whether a particular mutation actually affects mismatch repair. Conversely, the mutations in MSH2 detected in patients by genetic testing will be important tools in probing the mechanism of mismatch repair. This is a wonderful example of how basic science and clinical care can help each other. ■

Richard D. Kolodner is at the Ludwig Institute for Cancer Research, Department of Medicine and Cancer Center, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0660, USA.

e-mail: rkolodner@ucsd.edu

1. Lynch, H. T. & de la Chapelle, A. *J. Med. Genet.* **36**, 801–818 (1999).
2. Modrich, P. & Lahue, R. *Annu. Rev. Biochem.* **65**, 101–133 (1996).
3. Kolodner, R. & Marsischky, G. *Curr. Opin. Genet. Dev.* **9**, 89–96 (1999).
4. Obmolova, G., Ban, C., Hsieh, P. & Yang, W. *Nature* **407**, 703–710 (2000).
5. Lamers, M. H. *et al. Nature* **407**, 711–717 (2000).
6. Rayssiguier, C. *et al. Nature* **342**, 396–401 (1989).
7. Gong, J. G. *et al. Nature* **399**, 806–809 (1999).
8. Li, G.-M. *Oncol. Res.* **11**, 393–400 (1999).
9. Gradia, S. *et al. Mol. Cell* **3**, 255–261 (1999).
10. Marsischky, G. T. & Kolodner, R. D. *J. Biol. Chem.* **274**, 26668–26682 (1999).
11. <http://www.nfhdht.nl>

Evolutionary biology

Use it or lose it

Robert D. Holt

On page 736 of this issue Cooper and Lenski¹ tackle an enduring question in evolutionary biology — the origin and maintenance of ecological specialization^{2–4}. There are two different theories of how specialization can arise. Cooper and Lenski's aim was to test them using laboratory cultures of bacteria allowed to evolve over many generations.

First, however, step back and consider the striking fauna of caves, including fish and

other organisms in which the eyes are absent or greatly reduced (Fig. 1)^{5,6}. Blindness increases vulnerability to predation and decreases competitiveness in well-lit surface environments. So these species typically remain confined to caves.

Two processes can generate functional degeneration and habitat specialization of this sort. Mutations may arise that degrade the optical system but do not have other effects on the organism's function. In caves,



Figure 1 Eyeless example — the blind cave fish *Astyanax fasciatus mexicana*.

such mutations are neutral as far as natural selection is concerned, being neither beneficial nor harmful, and can drift to fixation (that is, become an enduring part of the genome). The gradual accumulation of neutral mutations could erode previously adaptive structures and indirectly lead to habitat specialization. Alternatively, natural selection may actively re-sculpt the organism, reallocating resources from eyes which are useless in the dark to more useful structures (a pattern of correlated effects that geneticists call 'antagonistic pleiotropy' and ecologists call 'trade-offs'). But it is difficult to tell the difference between these causes of specialization in most natural populations⁶. This is why Cooper and Lenski have explored the problem not in a cave, but in a laboratory study of microbial evolution.

The short generation time and high abundance of microbes, and the experimental control possible in lab conditions, make microbes good subjects for evolutionary study. Lenski and colleagues⁷ have been carrying out a long-term analysis of evolution in cultures of the bacterium *Escherichia coli*, and their system provides an ideal forum for investigating the genetic basis for a loss of ecological function — in this case, a decline in the ability of bacteria to use a variety of carbon sources for nutrition.

In their study¹, Cooper and Lenski confined replicate, genetically homogeneous lineages of bacteria to a simple environment (a minimal nutrient with glucose added). The lineages originated from an ancestral population sustained in a rich nutritive medium, with a smorgasbord of carbon substrates. Over 20,000 generations (taking about ten years in time), mutation resulted in increased adaptation of bacteria to glucose, with an accompanying decay in their ability to use alternative foodstuffs. In effect, *E. coli* growing on just glucose evolved to use a narrower diet, specializing to the available resource at the expense of their potential ability to live on alternative resources. This is comparable to cave fish which have lost a functional ability (sight) that is potentially useful in surface habitats.

Cooper and Lenski use several strands of evidence to argue that antagonistic pleiotropy has occurred. Mutations arise at random. So if functional loss occurs because of an accumulation of neutral mutations, one would expect an exponential decay in diet breadth. Instead, the decay was initially rapid but then slowed (mirroring the temporal dynamics of improved adaptation to glucose). Second, the pattern of functional loss was similar between the replicates. Because mutations arise independently in separate lineages, under the mutation-accumulation theory different functions should be knocked out first in different lineages. By and large, this did not happen.

Third, for one alternative resource

(ribose), the molecular mechanism coupling the improved use of glucose to the reduced use of the alternative is understood: a separate study showed that the mutations that lead to a loss of capacity on ribose provide a selective advantage on glucose. Finally, if the mere accumulation of mutations leads to reduced ability to use a varied diet, higher mutation rates should result in higher rates of loss of that ability. Fortunately, three lineages evolved changes that impaired their capacity for repairing damaged DNA, and so had much higher genome-wide mutation rates. But this elevated mutation rate had no significant effect on the rate of change in diet use.

Another, similar study⁸ has also examined evolution in a microbial population with high mutation rates, but with an experimental regime leading to severe 'bottlenecks'. This is the situation when a population is reduced to very few members, which enhances random processes of genetic evolution at the expense of selection, and in particular facilitates the fixation of deleterious mutations. In this study⁸, reduction in bacterial diet breadth was exponential in time, and unpredictable in pattern between replicate lineages — the patterns expected according to the mutation-accumulation theory. The contrast between this result and that of Cooper and Lenski¹ (where populations were large, greatly reducing the impact of random changes in genetic composition) also bolsters the case for trade-offs (antagonistic pleiotropy) being involved in ecological specialization in the latter experiment.

Nonetheless, many questions remain unanswered, both about Cooper and Lenski's experiments and their broader implications. A more quantitative characterization of the ancestral environment of the bacteria would be desirable, as would a mechanistic understanding of the metabolic constraints underlying antagonistic pleiotropy. Because Lenski's project is ongoing, further adaptive decay in diet breadth might emerge that is consistent with the mutation-accumulation hypothesis⁹. It would also be intriguing to examine evolutionary reversals. Can a lineage specialize to feed on a single nutrient and then re-evolve to use a generalized diet? If so, does this reversal become less likely, the longer the period of specialization? From comparative studies it seems that transitions from generalist to specialist occur more readily than in the other direction¹⁰. The lab system allows experimental assessment of such evolutionary asymmetries.

Microbes potentially provide solutions to environmental problems ranging from biological control of agricultural pests to cleaning up oil spills. So there are good practical reasons to understand the evolutionary constancy of microbial niches in natural environments, so as to assess the reliability and risks of such solutions. Generalizations

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

to natural populations should of course be made with caution, given the simplified conditions of lab cultures. The bacterial strains in Lenski and Cooper's cultures are strictly asexual, whereas most natural populations can sexually exchange genetic material. This matters if the availability of genetic variation is a rate-limiting factor in evolution. Sexuality could also alter the expression of antagonistic pleiotropy and the rate of evolution towards ecological specialization.

Moreover, evolution in this study was driven only by selection on the ability to live on abiotic resources in lab environments that are relatively homogeneous in space and time, and occupied by a single bacterial species. In natural communities, specialization occurs in heterogeneous arenas¹¹ seething with other species, including competitors, predators and parasites. Also, the functional degradations assessed by Lenski and Cooper reflect changes in relative ability to use resources, not absolute losses of functions. So these results may better predict the exclusion of a particular lineage through indirect competition for resources (where exclusion may emerge from slight differences in two strains' relative abilities to use resources), than population persistence in competitor-free habitats (where exclusion may require the nearly complete absence of function).

Despite these caveats, the study by Cooper and Lenski provides valuable insights into the evolutionary dynamics of ecological specialization. These dynamics are central to the evolution of species diversity¹², and understanding them may even shed light on evolution in the stygian depths occupied by blind cave organisms. ■

Robert D. Holt is at the Museum of Natural History and the Biodiversity Research Center, and the Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045, USA.

e-mail: predator@ukans.edu

- Cooper, V. S. & Lenski, R. E. *Nature* **407**, 736–739 (2000).
- Futuyama, D. J. & Moreno, G. *Annu. Rev. Ecol. Syst.* **19**, 207–234 (1988).
- Holt, R. D. *Evol. Ecol.* **10**, 1–11 (1996).
- Agrawal, A. A. *Ecology* **81**, 500–508 (2000).
- Culver, D. C. *Cave Life* (Harvard Univ. Press, 1982).
- Culver, D. C., Kane, T. C. & Fong, D. W. *Adaptation and Natural Selection in Caves* (Harvard, Univ. Press, 1995).
- Lenski, R. E. & Travisano, M. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814 (1994).
- Furchain, P. et al. *Genetics* **154**, 959–970 (2000).
- Kawecki, T. J., Barton, N. H. & Fry, J. D. *J. Evol. Biol.* **10**, 407–429 (1997).
- Kelley, S. T. & Farrell, B. D. *Evolution* **52**, 1731–1743 (1998).
- Kawecki, T. J. *Evol. Ecol.* **9**, 38–44 (1995).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).

Before striking gold in gold-ruby glass

The chemistry of the magic ingredient in this ancient glass is no longer a mystery.

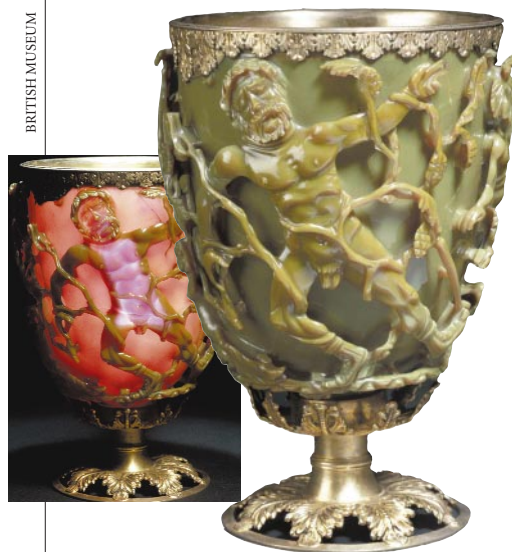


Figure 1 The Lycurgus Cup dates from Roman times. The glass appears green in daylight (reflected light), but red when light is transmitted from the inside of the vessel.

The red colour of gold-ruby glass is caused by small particles of metallic gold that form when the gold-containing colourless glass is annealed. But the chemical state of the gold before striking red has been a mystery — is it dissolved in the glass as individual neutral gold atoms, like a frozen-in metal vapour that precipitates on annealing¹, or as gold cations that must be reduced before the metallic clusters can grow and impart the lustrous colour^{2,3}? Here we use Mössbauer spectroscopy to show that this gold is monovalent in the colourless glass, forming linear bonds to two neighbouring oxygen atoms.

Gold has been used since ancient times to impart a red colour to glass — a fine example is the famous Lycurgus Cup in the British Museum⁴ (Fig. 1). Although described in Roman times, the production of gold-ruby glass was not rediscovered in Europe until the seventeenth century⁵, when sizeable red glass vessels were first made by adding 'Purple of Cassius', a precipitate of colloidal gold and stannic hydroxide, to the base glass.

Modern gold-ruby glass is usually made by adding gold to the raw materials as a solution of tetrachloroauric acid or potassium dicyanoaurate. At about 1,400 °C, the gold disperses in the melt, and the glass remains colourless after rapid cooling to ambient temperature. Only upon annealing between about 500–700 °C does the red colour strike as a result of the nucleation and growth of small metallic gold particles¹, whose optimum size for yielding a strong

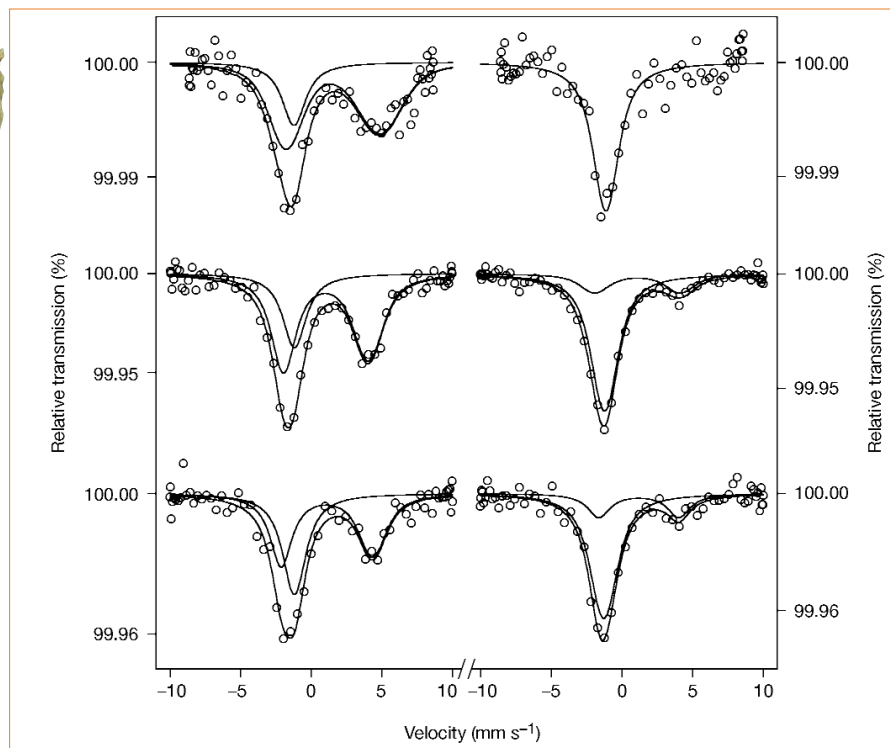


Figure 2 ¹⁹⁷Au Mössbauer spectra of quenched (left) and annealed (right) gold-ruby glasses measured at 4.2 K with a source of ¹⁹⁷Pt in ¹⁹⁶Pt metal. Glass 1 (top) is a commercial glass (from F. X. Nachtmann Bleikristallwerke GmbH) with the nominal composition (wt%) SiO₂, 46; PbO, 39; K₂O, 6; Na₂O, 2; B₂O₃, 1; As₂O₃, 1; Sb₂O₃, 1; SnO₂, 1, containing 200 p.p.m. Au introduced as HAuCl₄·xH₂O. It was melted at 1,450 °C and annealed at 520 °C for 5 h. Glass 2 (middle) has the nominal composition SiO₂, 74; Na₂O, 26, and contains 310 p.p.m. gold added as KAuC(N)₂. It was melted at 1,400 °C, quenched by pouring the melt into water, and annealed at 650 °C for 17 h. Glass 3 (bottom) has the nominal composition SiO₂, 70; Na₂O, 22; CaO, 8, and contains 200 p.p.m. gold introduced as HAuCl₄·xH₂O. It was melted at 1,400 °C, quenched by pouring onto a steel plate, and annealed at 590 °C for 10 h.

ruby colour is 5–60 nm (refs 2,6).

Figure 2 shows results we obtained from ¹⁹⁷Au Mössbauer spectroscopy for three samples of gold-ruby glass before and after annealing. ¹⁹⁷Au Mössbauer spectroscopy is useful for the *in situ* analysis of the state of gold in solids⁷. The technique is sensitive enough to pick up a low gold content and provides information on the oxidation state, coordination geometry and covalence of the gold.

'Glass 1' in Fig. 2 is a commercial glass with a high lead content, which limits the thickness of the Mössbauer absorbers to about 2 g cm⁻² because of the strong photoelectric absorption of the 77.3-keV γ-rays. The low gold content (200 p.p.m.) also makes the measurement of Mössbauer spectra difficult. We therefore also investigated samples of lead-free soda (glass 2) and soda-lime (glass 3) glasses. For these glasses, absorbers containing about 10 g cm⁻² could be used.

The typical single line of bulk metallic gold near -1.23 mm s⁻¹ is dominant in our Mössbauer spectra of the annealed

samples (Fig. 2, right traces), whereas the main component in the quenched samples is an asymmetrical quadrupole doublet (Fig. 2, left traces) consisting of two lines of equal area but with different widths and depths, which can be explained by a narrow distribution of isomer shifts and quadrupole splittings. The average isomer shifts (1.3 mm s⁻¹) and quadrupole splittings (6.5 mm s⁻¹) are typical of linearly coordinated monovalent gold and similar to those of compounds with a partially ionic character, such as [AuCl₂]⁻ or RAuCl, where R is an alkyl substituent⁷.

In the glass, the gold presumably bonds to two oxygen atoms: this idea is supported by the fact that the Mössbauer parameters of Au(I) in the quenched glasses are similar to those of ternary gold oxides such as CsAuO (ref. 8) or Na₃[AuO₂] (ref. 9), in which the gold is linearly coordinated to two oxygens. Au₂S, in which the Au(I) atoms are linearly coordinated to two sulphur atoms, has a similar Mössbauer spectrum¹⁰.

The Mössbauer spectra of the quenched glasses contain a minor component that is

attributable to metallic gold but which does not colour the glass; this probably represents large particles of bulk metallic gold that did not dissolve in the melt. The annealed lead-free glasses still contain minor amounts of monovalent gold because of incomplete precipitation. It has been proposed that the gold is first reduced to individual Au⁰ atoms and then precipitates by nucleation and growth¹¹, but our spectra show no feature that could be attributed to individual Au⁰ atoms, although this may be because only the quenched and largely precipitated states have been studied so far.

The presence of monovalent gold in quenched samples obtained from both HAuCl₄·xH₂O and KA(CN)₂ shows that the oxidation state of the gold in the glass does not depend on the oxidation state of the gold in the starting material. Remelting of a coloured, annealed specimen of glass 2 at 1,400 °C and subsequent quenching again produced a colourless glass, whose Mössbauer spectrum showed that the metallic gold had transformed back to Au(I). This shows that the transformation of Au(I) to metallic gold is reversible and that monovalent gold is stabilized by the glass matrix.

Social insects

Facultative worker policing in a wasp

Kin-selection theory predicts that in social-insect colonies where the queen has mated multiple times, the workers will enforce cooperation by policing each other's reproduction^{1–4}. We have discovered a species, the wasp *Dolichovespula saxonica*,



Figure 1 Worker egg-laying in *Dolichovespula saxonica*. Workers police other workers' reproduction more in colonies where the queen has mated multiple times than in those where she has mated only once.

F. E. Wagner*, S. Haslbeck*, L. Stievano†, S. Calogero†, Q. A. Pankhurst‡, K.-P. Martinek§

*Physik-Department E15, Technische Universität München, 85748 Garching, Germany

†Dipartimento di Chimica Fisica, Università "Ca' Foscari", Calle Larga S. Marta 2137, 30123 Venice, Italy

‡Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

§F. X. Nachtmann Bleikristallwerke GmbH, 94566 Riedelshütte, Germany

1. Weyl, W. A. *Coloured Glasses* 380–400 (Society of Glass Technology, Sheffield, 1951).
2. Badger, A. E., Weyl, W. & Rudow, H. *Glass Ind.* **20**, 407–414 (1939).
3. Stookey, S. D. *J. Am. Ceram. Soc.* **32**, 246–249 (1949).
4. Barber, D. J. & Freestone, I. C. *Archaeometry* **32**, 33–45 (1990).
5. Frank, S. *Glass Technol.* **25**, 47–50 (1984).
6. Williams, J. A., Rindone, G. E. & McKinstry, H. A. *J. Am. Ceram. Soc.* **64**, 709–713 (1981).
7. Parish, R. V. in *Mössbauer Spectroscopy Applied to Inorganic Chemistry* Vol 1 (ed. Long, G. J.) 577–617 (Plenum, New York, 1984).
8. Wasel-Nielsen, H.-D. & Hoppe, R. Z. *Anorg. Allg. Chem.* **359**, 36–40 (1968).
9. Wagner, G. & Hoppe, R. Z. *Anorg. Allg. Chem.* **549**, 26–34 (1987).
10. Faltens, M. O. & Shirley, D. A. *J. Chem. Phys.* **53**, 4249–4264 (1970).
11. Rao, P. & Doremus, R. J. *Non-Cryst. Solids* **203**, 202–205 (1996).

in which some queens mate once and others mate many times, and in which workers frequently attempt reproduction, allowing this prediction to be tested directly. We find that multiple mating by the queen leads to mutual policing by workers, whereas single mating does not.

Workers in most species of social Hymenoptera (bees, ants and wasps) cannot mate but can produce unfertilized, male eggs. Workers and the queen therefore compete over male production. If the queen mates only once, workers are more closely related to the sons of other workers ($r=0.375$) than to those of their mother queen ($r=0.25$) and, in conflict with the queen, should prefer to rear other workers' sons. But if the queen mates multiple times, workers are more related to the queen's sons than to other workers' sons. This is expected to lead to worker policing, where workers attempt to stop each other from reproducing^{1–3}.

Mutual policing by egg eating occurs in the honeybee *Apis mellifera*⁵ and in the common wasp *Vespa vulgaris* (K.R.F. and F.L.W.R., manuscript submitted), whose queens are multiply mated, but not in the stingless bees⁶ and bumblebees⁷, whose queens typically mate only once. A direct within-species test is critical, however, because these taxa differ in many ways apart from kinship and, with only a few independent data points, the trend is not statistically significant.

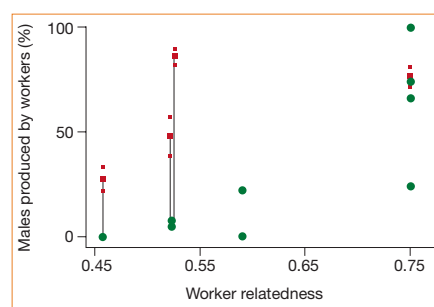


Figure 2 Pedigree relatedness among workers versus adult male production by workers in nine colonies of the wasp *Dolichovespula saxonica* (green circles). Estimates of the percentage of the male eggs that are laid by workers in four of the colonies are also shown (red squares). Nests were collected from the New Forest, UK, in 1999. Pedigree relatedness was estimated by inspection of 20 worker genotypes from each nest at three DNA microsatellite loci (Rufa 5, 13, 18)⁸ to determine the number of fathers and their relative paternity. Male production by workers was estimated by genotyping 17–30 males from each nest at all informative loci. The percentage of worker-produced adult males detected for each nest was adjusted according to the probability of detecting workers' sons $\sum_i p_i(1-0.5^i)$, where n is the number of patriline in the nest, p_i is the proportional representation of the i th patriline and i is the number of informative loci analysed at the i th patriline. This calculates the probability that a worker-produced male will inherit an allele from the queen's mate, thereby making him distinguishable from the queen's sons. The percentage of male eggs laid by workers is based on 80 h observation of four free-flying nests housed in glass-sided boxes at the Institute of Terrestrial Ecology, Furzebrook, Dorset, UK (101 queen-laid and 164 worker-laid eggs in total). Because the precise sex ratio of queen-laid eggs is unknown, three points (red squares) corresponding to a range of estimates, 2M:1F (bottom), 1M:1F (middle) and 1M:2F (top), are shown. The first produces the lowest estimate of worker reproduction and is highly conservative as *Dolichovespula* rear about twice as many females as males so that the queen probably lays more female eggs than male. The estimates of the percentage of male eggs laid by workers for the singly mated colony ($r=0.75$) lie on top of the corresponding adult male production data.

In the vespine wasp *Dolichovespula saxonica* (Fig. 1), the single queen may be either singly or multiply mated, leading to societies that differ only in kinship. We used DNA microsatellites⁸ to analyse worker relatedness and male production in nine colonies.

There was a strong positive correlation between worker relatedness and male production by workers (Fig. 2; Spearman's rank correlation, $P<0.004$). Observation revealed that this was caused not by differences in worker laying — workers in colonies with singly and multiply mated queens laid nearly identical proportions of male eggs (χ^2 , $P>0.86$, for all three estimates of queen-laid sex ratio) — but by removal of worker-laid eggs in the nests with multiply mated queens. In the observation colonies, worker egg production was significantly different from adult male production in the colonies of multiply mated queens (χ^2 , $P<0.001$), but not in those of singly mated queens (χ^2 , $P>0.39$, for all

three estimates of queen-laid sex ratio) (Fig. 2). All eight nests sampled contained reproductive workers, with 1–4 of the 20 workers examined in each nest possessing full-size eggs, also indicating that the amount of worker laying is comparable in all colonies. Interestingly, the relatedness at which policing occurs is slightly higher than that predicted from relatedness alone, suggesting that costs associated with worker male production might also favour worker policing².

To our knowledge, our results are the first direct evidence that multiple mating of queens causes mutual policing by workers. Policing favours workers' cooperation by preventing their reproduction^{2–4}, but as not all *D. saxonica* colonies have policing, here it is only partially effective in countering worker reproduction. Worker policing and cooperation have progressed further in the honeybee *Apis mellifera*, which has policing in all colonies, negligible worker reproduction and large colonies⁵. *D. saxonica* may represent an intermediate stage in the evolution of enforced cooperation. By discouraging policing in some colonies, the close kinship caused by single mating of the queen may paradoxically retard social evolution in this species and others like it.

Kevin R. Foster,

Francis L. W. Ratnieks

Laboratory of Apiculture and Social Insects,
Department of Animal and Plant Sciences, Sheffield
University, Western Bank, Sheffield S10 2TN, UK
e-mail: Bop97krf@shef.ac.uk

1. Starr, C. K. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 427–464 (Academic, Orlando, 1984).
2. Ratnieks, F. L. W. *Am. Nat.* 132, 217–236 (1988).
3. Frank, S. A. *Nature* 377, 520–522 (1995).
4. Keller, L. (ed.) *Levels of Selection in Evolution* (Princeton Univ. Press, Princeton, NJ, 1999).
5. Ratnieks, F. L. W. & Visscher, P. K. *Nature* 342, 796–797 (1989).
6. Peters, J. M., Queller, D. C., Imperatriz-Fonseca, V. C., Roubik, D. W. & Strassmann, J. E. *Proc. R. Soc. Lond. B* 266, 379–384 (1999).
7. Estoup, A., Scholl, A., Pouvreau, A. & Solignac, M. *Mol. Ecol.* 4, 89–93 (1995).
8. Thorén, P. A., Paxton, R. J. & Estoup, A. *Ins. Mol. Biol.* 4, 141–148 (1995).
9. Foster, K. R., Ratnieks, F. L. W. & Raybould, A. F. *Mol. Ecol.* 9, 735–742 (2000).

Embryonic development

Maternal effect of *Hsf1* on reproductive success

A protein known as heat-shock factor-1 (HSF1) is a major transactivator of stress-inducible genes in response to environmental changes, but it is also implicated in extra-embryonic development and female fertility in mice^{1,2}. Here we show that mouse embryos whose mothers lack this protein are unable to develop properly beyond the zygotic stage, although oocytes were ovulated and fertilized normally. Wild-

type spermatozoa do not save zygotes from lethality, indicating that the reproductive failure of females deficient in this factor is caused by a 'maternal effect' mutation³, and that HSF1 from the mother normally controls early post-fertilization development.

Female mice lacking the gene encoding HSF1 (*Hsf1*^{−/−} females) have normal ovaries and reproductive tracts, indicating that folliculogenesis and oogenesis do not require HSF1 expression, by contrast with the *Drosophila Hsf* mutation⁴ and others affecting fertility in mammals⁵. Ovulated eggs of *Hsf1*^{−/−} females remain properly arrested at metaphase II until fertilization, after which zygotes form with two pronuclei and a second polar body (Fig. 1a). Although mutant embryos produced by *Hsf1*^{−/−} females can initiate early development, they do not survive after transplantation into the oviduct of *Hsf1* wild-type mice, indicating that the causes of *Hsf1*^{−/−} infertility are intrinsic rather than extrinsic (see <http://cardiology.swmed.edu/maternalhsf1/default.htm>).

To determine at which stage development breaks down, we analysed the cleavage efficiency of preimplantation embryos produced by *Hsf1*^{−/−} females, using *Hsf1*^{+/+} females as controls, mated with either wild-type or homozygous males (Fig. 1a–d). The percentage of blastocysts at 3.5 days post-coitum (d.p.c.) was the same and independent of the genotype of the father (92.5 versus 91%, respectively; Fig. 1i).

In contrast, the embryos of homozygous females mated with either *Hsf1*^{+/+} or *Hsf1*^{−/−} males were blocked mainly at the 1-cell stage and development at 3.5 d.p.c. was poor (Fig. 1j). Unlike the embryos produced by heterozygous females, none of these mutant embryos reached the blastocyst stage at the appropriate time, and only 6 of 33 (18%) embryos from the *Hsf1*^{−/−} × *Hsf1*^{−/−} intercross reached the 2-cell stage. A significantly higher fraction of those sired by *Hsf1*^{+/+} males progressed beyond the 1-cell division (20/34; 58%) ($P < 0.001$), but the difference does not persist beyond this stage (Fig. 1j). Our findings show that the survival of the offspring is determined mainly by the maternal genotype, and that *Hsf1*^{−/−} is therefore a maternal-effect mutation³.

Embryos might die at the 1–2-cell stage because of a failure to initiate the zygotic transcriptional activity required to complete the transition from maternal to embryonic control of development⁶. As spontaneous expression of Hsp70.1 indicates the onset of transcriptional activity at the late 1-cell stage^{7–9}, we used Hsp70.1-Luc transgenic lines, in which a luciferase reporter gene is driven by the murine Hsp70.1 promoter⁸, to test this idea directly. At 1.5 d.p.c., embryos from *Hsf1*^{+/+} and *Hsf1*^{−/−} females that had been mated with

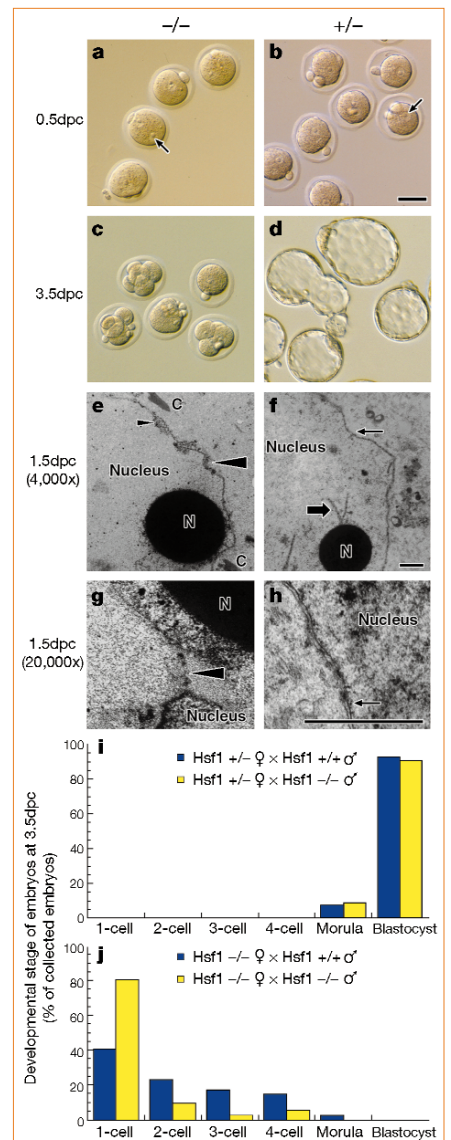


Figure 1 Developmental fates of preimplantation embryos collected from *Hsf1*^{−/−} and *Hsf1*^{+/+} females. **a, b**, One-cell embryos have similar pronucleus-like structures (arrows) at 0.5 d.p.c. **c, d**, At 3.5 d.p.c., when almost all embryos from matings between *Hsf1*^{+/+} females and either *Hsf1*^{+/+} or *Hsf1*^{−/−} males produce blastocysts, those from *Hsf1*^{−/−} females are blocked or unequally cleaved. Scale bar, 50 μ m. **e–h**, Reduced survival of mutant embryos corresponds with ultrastructural abnormalities in 2-cell embryos: heterochromatin (small arrowhead in **e**), non-specific crystalloids ('C' in **e**), absence of intranuclear membrane disruption in nuclear membranes (arrowheads in **e** and **g** versus small arrows in **f** and **h**). Scale bars, 1 μ m; N, nucleolus. **i, j**, Developmental profiles for *Hsf1*^{+/+} (**i**) and *Hsf1*^{−/−} (**j**) females.

Hsf1^{+/+} transgenic males both showed luciferase activity significantly above the background level (50 ± 0.5 ALUs versus 41 ± 0.4 ALUs, respectively) (<http://cardiology.swmed.edu/maternalhsf1/default.htm>). The *Hsf1*-mutant zygotes therefore apparently express heat-shock proteins, showing that zygotic transcriptional activity can begin without HSF1.

Our results do not exclude the possibility that the level of transcriptional activity, or the control of the target genes to be tran-

scribed¹⁰, might have been altered in mutant embryos^{11,12}. Furthermore, ultrastructural abnormalities affect mainly the nuclei of embryos from null females at the 2-cell stage, indicating that, even with transcriptional activity, reduced survival is in part related to the decrease in structural integrity of the early embryonic nucleus (Fig. 1e–h).

The key role for control by maternal-origin HSF1 in early development raises the possibility that in defective form it may be a cause of post-fertilization abnormalities associated with infertility in mammals, including humans.

E. Christians*, **A. A. Davis†‡**,
S. D. Thomas†, **I. J. Benjamin †§**

*Department of Histology and Embryology, Faculty of Veterinary Medicine, University of Liège, 20 Boulevard de Colonster, 4000 Liège, Belgium

†Department of Internal Medicine and §Division of Cell and Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-8573, USA

e-mail: ivor.benjamin@utsouthwestern.edu

‡Present address: Alcon Laboratories, Fort Worth, Texas 76134-2099, USA

- McMillan, D. R., Xiao, X., Shao, L., Graves, K. & Benjamin, I. J. *J. Biol. Chem.* **273**, 7523–7528 (1998).
- Xiao, X. *et al.* *EMBO J.* **18**, 5943–5952 (1999).
- Schupbach, T. & Wieschaus, E. *Genetics* **121**, 101–117 (1989).
- Jedlicka, P., Mortin, M. A. & Wu, C. *EMBO J.* **16**, 2452–2462 (1997).
- Greenhouse, S., Rankin, T. & Dean, J. *Am. J. Hum. Genet.* **62**, 1282–1287 (1998).
- Schultz, R. M. *BioEssays* **15**, 531–538 (1993).
- Bensaude, O., Babinet, C., Morange, M. & Jacob, F. *Nature* **305**, 331–333 (1983).
- Christians, E., Campion, E., Thompson, E. M. & Renard, J. P. *Development* **121**, 113–122 (1995).
- Christians, E. *et al.* *Mol. Cell Biol.* **17**, 778–788 (1997).
- Bevilacqua, A., Fiorenza, M. T. & Mangia, F. *Development* **127**, 1541–1551 (2000).
- Latham, K. E., Garrels, J. I., Chang, C. & Solter, D. *Development* **112**, 921–932 (1991).
- Davis, W. Jr & Schultz, R. M. *Dev. Biol.* **218**, 275–283 (2000).

Biodiversity

Coexistence and resource competition

How large numbers of species coexist on a seemingly limited number of different resources is a classic problem in ecology¹, and attempts have been made to solve it experimentally. But we are not convinced that Huisman and Weissing's² proposal to add non-stationary dynamics in species abundance to the list of possible explanations offers any new insight into this biodiversity enigma.

Early ideas were based on competitive exclusion, in which two (or more) species compete for shared resources but only the more proficient survives³, and the number of coexisting species, n , does not exceed the

number of available resources, k . But these ideas based on competitive exclusion cannot explain how rich ensembles of species can coexist on a limited number of resources (that is, $n > k$), as in aquatic ecosystems.

A solution to this biodiversity problem has been suggested by Huisman and Weissing² in the form of a model demonstrating the possibility of coexistence of two or more competitors under conditions of fluctuating (periodic or chaotic) community dynamics, a finding central to contemporary ecological theory⁴. However, we believe that their solution to the $n > k$ puzzle has long been known^{5–13}.

It was first inferred from a computer simulation over 25 years ago⁵ that two species can coexist on one biotic resource, with coexistence occurring along what appeared to be a periodic orbit, as was also shown by Huisman and Weissing². This was confirmed analytically and expanded to cover the more general case of n species coexisting on one biotic resource⁹. The fact that coexistence depends on the nonlinearity of the species-specific growth functions and on the lack of system equilibrium — the central issue discussed by Huisman and Weissing² — has also been demonstrated earlier⁸. The plankton paradox (whereby $n > k$) was thus resolved about two decades after it was first posed as a biodiversity problem.

Per Lundberg*, **Esa Ranta†**, **Veijo Kaitala†‡**,
Niclas Jonzén*

*Department of Theoretical Ecology, Ecology Building, Lund University, 223 62 Lund, Sweden
e-mail: per.lundberg@teorekol.lu.se

†Integrative Ecology Unit, Division of Population Biology, Department of Ecology and Systematics, PO Box 17 (Arkadiankatu 7), 00014 University of Helsinki, Finland

‡Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, 40351 Jyväskylä, Finland

- Hutchinson, G. E. *Am. Nat.* **95**, 137–145 (1961).
- Huisman, J. & Weissing, F. J. *Nature* **402**, 407–410 (1999).
- Gause, G. J. *The Struggle for Existence* (Williams & Wilkins, Baltimore, 1934).
- Sommer, U. *Nature* **402**, 366–367 (1999).
- Koch, A. L. *J. Theor. Biol.* **44**, 387–395 (1974).
- Armstrong, R. A. & McGehee, R. J. *Theor. Biol.* **56**, 499–502 (1976).
- Armstrong, R. A. & McGehee, R. *Theor. Popul. Biol.* **9**, 317–328 (1976).
- Armstrong, R. A. & McGehee, R. *Am. Nat.* **115**, 151–170 (1980).
- McGehee, R. & Armstrong, R. A. *J. Diff. Equ.* **23**, 30–52 (1977).
- Kalan, J. L. & Yorke, J. A. *Am. Nat.* **111**, 1032–1036 (1977).
- Ebenhöh, W. *Ecol. Modell.* **75/76**, 83–98 (1994).
- Yodanis, P. *Introduction to Theoretical Ecology* (Harper & Row, New York, 1989).
- Fryxell, J. M. & Lundberg, P. *Individual Behavior and Community Dynamics* (Chapman & Hall, London, 1997).

briEF COMMUNICATIONS is intended to provide a forum for short, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com) or on request from nature@nature.com

Huisman and Weissing reply — How can we explain the biodiversity of rainforests and coral reefs if we do not understand the species diversity of phytoplankton in a droplet of water¹? In our attempt² to solve the plankton paradox, we showed that multispecies competition may generate oscillations and chaos and that these fluctuations create opportunities for the coexistence of many species. Lundberg *et al.* question the novelty of our findings and argue that the plankton paradox was resolved 25 years ago.

The earlier findings^{3–5} quoted by Lundberg *et al.* are interesting because they show that multiple species can coexist on a single biotic resource. However, biotic resources are irrelevant to an explanation of the plankton paradox. This paradox is concerned with phytoplankton¹, and phytoplankton species do not compete for biotic resources — they use abiotic resources such as nitrogen, phosphorus, silicon, iron and light.

Also, the mechanisms that generate the fluctuations differ. The models to which Lundberg *et al.* refer consider predators competing for biotic prey. It is well known that predator–prey interactions easily generate fluctuations⁶. The fluctuations that we investigated, however, are not predator–prey oscillations. We discovered that models of competition for abiotic resources^{2,7} may also generate fluctuations, and that these competitive fluctuations allow a high species diversity. We have therefore shown that competition itself can generate the fluctuations that favour species coexistence. This resolves the plankton paradox.

Apart from these differences with the earlier findings^{3–5} in terms of the type of resource and the source of the fluctuations, there is an important similarity, as pointed out by Lundberg *et al.* All of these studies indicate that non-equilibrium dynamics generated by species interactions can have a major impact on biodiversity.

Jef Huisman*, **Franz J. Weissing†**

*IBED–Microbiology, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands

e-mail: jef.huisman@chem.uva.nl

†Department of Genetics, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

- Hutchinson, G. E. *Am. Nat.* **95**, 137–145 (1961).
- Huisman, J. & Weissing, F. J. *Nature* **402**, 407–410 (1999).
- Koch, A. L. *J. Theor. Biol.* **44**, 387–395 (1974).
- Armstrong, R. A. & McGehee, R. J. *Theor. Biol.* **56**, 499–502 (1976).
- Armstrong, R. A. & McGehee, R. *Am. Nat.* **115**, 151–170 (1980).
- Volterra, V. *Nature* **118**, 558–560 (1928).
- Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).

A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization

Philip W. Boyd¹, Andrew J. Watson², Cliff S. Law³, Edward R. Abraham⁴, Thomas Trull⁵, Rob Murdoch⁴, Dorothee C. E. Bakker², Andrew R. Bowie^{6,3}, K. O. Buesseler⁷, Hoe Chang⁴, Matthew Charette⁷, Peter Croot⁸, Ken Downing⁴, Russell Frew⁹, Mark Gall¹⁰, Mark Hadfield⁴, Julie Hall¹¹, Mike Harvey⁴, Greg Jameson³, Julie LaRoche¹², Malcolm Liddicoat³, Roger Ling³, Maria T. Maldonado^{13,14}, R. Michael McKay¹⁵, Scott Nodder⁴, Stu Pickmere¹¹, Rick Pridmore¹¹, Steve Rintoul¹⁶, Karl Safi¹¹, Philip Sutton⁴, Robert Strzepek¹⁷, Kim Tanneberger², Suzanne Turner², Anya Waite¹⁸ & John Zeldis¹⁰

¹ National Institute of Water and Atmosphere, Centre for Chemical and Physical Oceanography, Department of Chemistry, University of Otago, Dunedin, New Zealand; ² School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK; ³ Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, Devon PL1 3DH, UK; ⁴ National Institute of Water and Atmosphere, Greta Point, PO Box 14-901, Wellington, New Zealand; ⁵ Antarctic Co-operative Research Centre, University of Tasmania, GPO Box 252-80, Hobart, Tasmania, 7001, Australia; ⁶ Department of Environmental Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK; ⁷ Department of Marine Chemistry and Geochemistry MS25, Woods Hole Oceanographic Institute, Woods Hole, Massachusetts 02543, USA; ⁸ Netherlands Institute for Sea Research (NIOZ), Department of Marine Chemistry and Geology, Postbus 59, 1790 AB Den Burg – Texel, The Netherlands; ⁹ Department of Chemistry, University of Otago, PO Box 56, Dunedin, New Zealand; ¹⁰ National Institute of Water and Atmosphere, Christchurch, PO Box 8602, Christchurch, New Zealand; ¹¹ National Institute of Water and Atmosphere, Hamilton, Box 11-115, Hamilton, New Zealand; ¹² Institut fuer Meereskunde, Universitaet Kiel, Duernsternbrooker Weg 20 D-24105 Kiel, Germany; ¹³ Biology Department, McGill University, 1205 Avenue, Dr. Penfield, Montreal, PQ H2T 2V8, Canada; ¹⁴ Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio 43403, USA; ¹⁵ CSIRO Division of Marine Research, GPO Box 1538, Hobart, Tasmania, 7001, Australia; ¹⁶ University of British Columbia, Departments of Botany and Oceanography, 6270 University Boulevard, Vancouver, BC, Canada, V6T 1Z4; ¹⁷ Centre for Water Research, Department of Environmental Engineering, University of Western Australia, Nedlands 6907, Western Australia, Australia.

Changes in iron supply to oceanic plankton are thought to have a significant effect on concentrations of atmospheric carbon dioxide by altering rates of carbon sequestration, a theory known as the 'iron hypothesis'. For this reason, it is important to understand the response of pelagic biota to increased iron supply. Here we report the results of a mesoscale iron fertilization experiment in the polar Southern Ocean, where the potential to sequester iron-elevated algal carbon is probably greatest. Increased iron supply led to elevated phytoplankton biomass and rates of photosynthesis in surface waters, causing a large drawdown of carbon dioxide and macronutrients, and elevated dimethyl sulphide levels after 13 days. This drawdown was mostly due to the proliferation of diatom stocks. But downward export of biogenic carbon was not increased. Moreover, satellite observations of this massive bloom 30 days later, suggest that a sufficient proportion of the added iron was retained in surface waters. Our findings demonstrate that iron supply controls phytoplankton growth and community composition during summer in these polar Southern Ocean waters, but the fate of algal carbon remains unknown and depends on the interplay between the processes controlling export, remineralisation and timescales of water mass subduction.

The supply of iron, via dust, to the ocean in the past is thought to have differed significantly from that of the present day¹, and may alter again in the near future owing to the additional influence of anthropogenic changes in aridity and hence dust flux². Such variations in iron supply are thought to alter the magnitude of oceanic primary² and export¹ production and the subsequent uptake of atmospheric CO₂. Mesoscale iron-enrichment of the ocean (Ironex II, 3.5°S, 105°W)³ showed that iron supply controls phytoplankton processes⁴ in the high-nitrate low-chlorophyll (HNLC) equatorial Pacific Ocean, resulting in enhanced algal stocks³ and associated macronutrient uptake³ and CO₂ drawdown⁵. However, modelling simulations of the effect of iron fertilization on equatorial Pacific waters indicate that it is unlikely to be a significant region for iron-mediated carbon sequestration⁶. In contrast, the Southern Ocean is the largest repository of unused macronutrients in surface waters⁵, and an important region for intermediate- and deep-water formation; consequently, it is a potential site for enhanced sequestration of carbon^{5,6}.

The Southern Ocean is thought to have a disproportionate influence on the functioning of the global carbon cycle in both the present⁷ and the geological past⁸. Indeed, palaeoceanographic proxies suggest that export production during the Last Glacial

Maximum was increased in these waters, probably due to increased iron supply⁹. Today, much of this region is characterized by low iron levels and/or deep mixed layers¹⁰ where iron^{10,11}, light^{12,13} or iron/light co-limitation¹⁴ may control phytoplankton growth rates. Silicate availability, particularly during summer, may determine diatom growth rates (V. Franck, personal communication) and nothing is known about the response of large grazers¹⁵ to increased algal stocks. These issues were addressed by conducting the *in situ* mesoscale Southern Ocean iron-release experiment (SOIREE) over 13 days in February 1999 (Fig. 1a). SOIREE confirms that iron supply controls phytoplankton processes at this site in summer, but as no significant iron-increased export production was observed we could not test the second tenet of the 'iron hypothesis'¹, which requires that fixed carbon is then sequestered in the deep ocean.

Site selection and survey

The Southern Ocean is characterized by several circumpolar fronts, which separate regions of relatively uniform water-mass properties¹⁶ and coincide with strong current cores of the Antarctic Circumpolar Current (ACC)¹⁷. Mixed-layer depth generally decreases (Fig. 1a), and macro-nutrient levels increase¹⁸, from north to south in this region. The SOIREE site (61°S, 140°E) was chosen to be representative of the properties (such as mixed-layer depth, Fig. 1a) of a broad region of circumpolar HNLC waters, yet have sufficiently low

¹⁴ Present address: School of Marine Sciences, University of Maine, Orono, Maine 04469, USA.

horizontal current shear to maximize the time that fertilized waters could be tracked. Before SOIREE we made a transect about 115 km northwards on the 139° E meridian (that is, upstream of the likely site) and about 30 km east to west, to constrain spatial variability and finalize site selection. Hydrographic and expendable bathy-thermograph (XBT) results indicated a buoyant seasonal mixed layer of about 65 m depth (Fig. 1b), as suggested by climatological

data sets (Fig. 1a), and clear zonal similarity to hydrographic¹⁹ and XBT²⁰ sections.

Mixed-layer nitrate ($\sim 25 \pm 1 \mu\text{M}$)¹⁸ and phosphate ($\sim 1.5 \pm 0.2 \mu\text{M}$) levels were relatively high, silicate ($\sim 10 \pm 0.4 \mu\text{M}$)¹⁸ was intermediate, and dissolved iron levels were low ($\sim 0.08 \pm 0.03 \text{ nM}$)¹⁰, for polar waters. Chlorophyll *a* concentration was 0.25 mg m^{-3} , and dominated by pico-eukaryotes ($\sim 50\%$ chlorophyll *a*, see Methods),

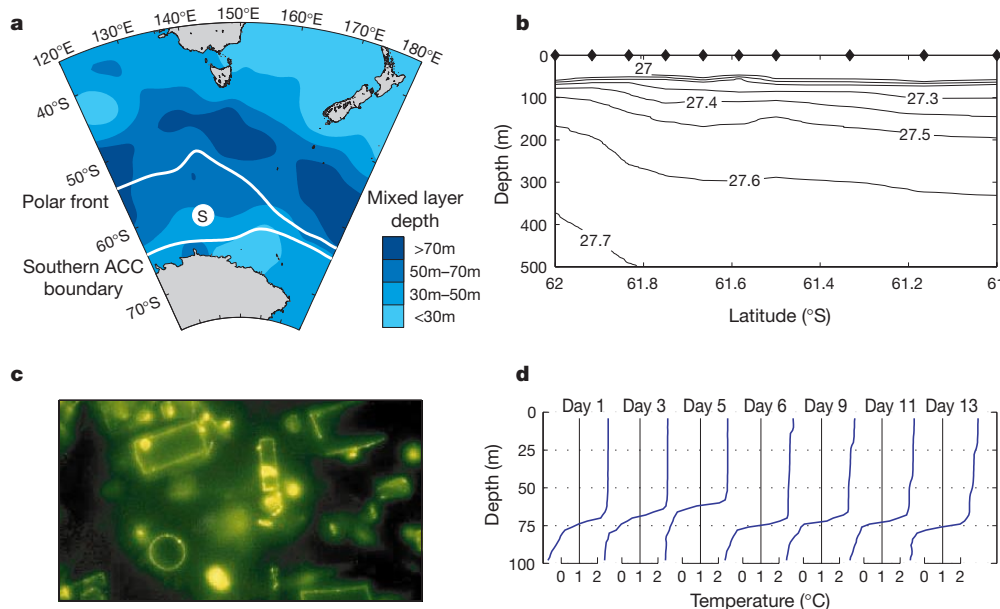


Figure 1 Site selection, pre-release survey and water-column structure. **a**, Summer mean mixed-layer depth—calculated from the World Ocean Atlas database—for the Australian–Pacific sector of the Southern Ocean. The location of the experiment (letter S), and the positions of the main circumpolar features are shown. ACC, Antarctic Circumpolar Current. **b**, Upper-ocean density-anomaly section (contours, kg m^{-3}) along the 139° E meridian during the survey before SOIREE. Contouring is based on 10 CTD casts (filled

diamonds). **c**, Photomicrograph of diatoms (width of image, 350 μm) from surrounding HNLC waters (10 m depth) on 12 February 1999. The green fluorescence indicates iron-stressed diatoms (flavodoxin immunofluorescence assay²¹). **d**, Temporal evolution of vertical temperature structure within the patch (surrounding HNLC waters also exhibited these trends).

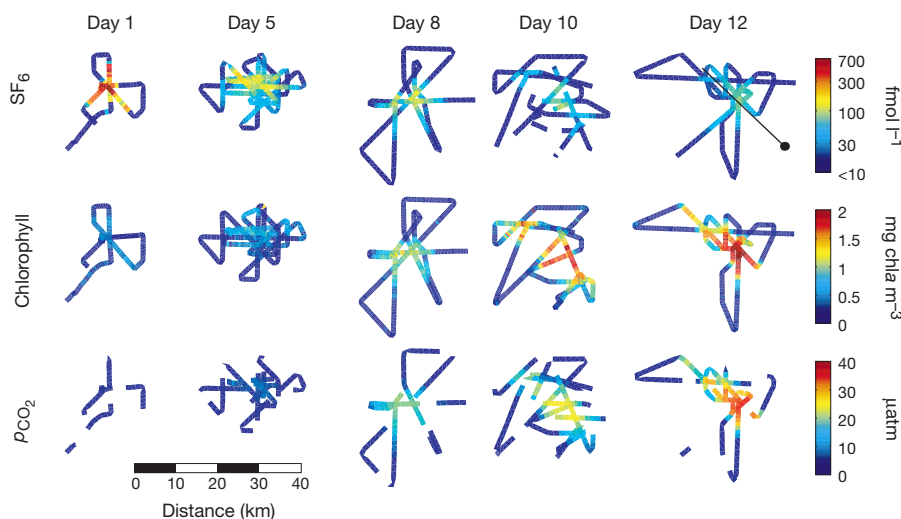


Figure 2 Temporal evolution of physical, chemical and biological properties during SOIREE. The figure shows maps of subsurface (5 m depth) SF₆ concentrations (top row), chlorophyll *a* concentrations (middle row), and changes in partial pressure of CO₂ (p_{CO_2} ; bottom row) within the iron-fertilized waters, relative to outside the patch. Taking into account changes in the area of the patch, the upper limit of spatial variability of SF₆, chlorophyll *a* and p_{CO_2} levels within the patch was estimated to <15%, <20% and <10% of the daily mean value, respectively. The sensitivity of the SF₆ analytical system was ten times lower than the background signal, the precision and limit of detection of the

chlorophyll *a* method was 0.04 and 0.01 mg m^{-3} , respectively, and both the accuracy and precision of the p_{CO_2} measurements was $\sim 1 \mu\text{atm}$. The 'ribbon' plots denote the vessel track during the daily SF₆ surveys, and the expanding patch size with time. The solid line in the SF₆ plot (day 12) is the location of the southwest–northeast CTD transect through the patch on 22 February (day 13) shown in Fig. 4. The patch moved no more than $\sim 10 \text{ km d}^{-1}$ to the northeast, then later to the southeast, and the patch area (based on a linear fit) increased by $\sim 15 \text{ km}^2 \text{ d}^{-1}$.

with an 'inoculum' of diatoms ($\sim 20\%$ chlorophyll *a*). Cells exhibited low values of photosynthetic competence ($F_v/F_m = 0.22$) and the resident large diatoms contained flavodoxin, a protein present under iron-stressed conditions²¹ (Fig. 1c), both suggesting iron-limitation. Thus, we had favourable conditions for determining a

response to iron-enrichment. We started SOIREE south of the polar front and north of the Southern ACC front (Fig. 1a), in a 65-m mixed layer (Fig. 1d) on 9 February 1999. Occupation of the site ended on 22 February (day 13). Times are in UTC (that is, GMT), and day 1 is defined as the 24-h period starting at 12:00 UTC, 9 February.

The experiment

Surface waters were fertilized with iron following an approach²²—employed in Ironex I²³ and II³—that added a conservative tracer sulphur hexafluoride (SF_6)²⁴ concurrently to provide rapid identification of iron-enriched waters. The release of iron and SF_6 over an area of $\sim 50 \text{ km}^2$ increased levels of SF_6 and dissolved iron to 300 fmol l^{-1} (mean) and $\sim 3 \text{ nM}$, respectively (Figs 2 and 3a). Daily mapping of the spatial extent of labelled waters (that is, the patch) indicated that they had mixed in—within 48 h—to cover 100 km^2 , increasing to $\sim 200 \text{ km}^2$ by day 13. SF_6 levels remained greater than 50 fmol l^{-1} throughout SOIREE (Fig. 2), and thus no further SF_6 was added. Interspersed between mapping, hydrographic stations within (at the patch centre) and outside the patch indicated little change in the physical structure of the upper ocean during days 1–4. Thus, trends in data obtained when the ship was underway (Fig. 2) were applicable over the upper 65 m. On days 5/6 and 8/9, relatively calm conditions led to the development of transient thermal structures within the seasonal mixed layer, both inside (Fig. 1d) and outside the patch. These features were eroded and mixed down to 40 m on days 10/11. Subsequent warming resulted in three temporary strata between 0 and 20 m, 20 and 50 m, and 50 and 65 m, within the seasonal mixed layer on day 13.

Daily surveys of dissolved iron revealed that after iron infusion no. 1, levels were initially increased (Fig. 3a), similar to those in the iron-rich regions of the Atlantic polar front¹¹, and comparable to concentrations during Ironex I²³ and II³. By day 2 levels of dissolved iron decreased rapidly, probably due both to the patch spreading, and to its conversion to particulate form, as total (unfiltered) iron levels remained at about 2 nM. Dissolved iron levels were used as the criterion to determine when to add more iron during SOIREE; three further infusions (nos 2, 3 and 4) were required on days 3, 5 and 7, respectively, when levels of dissolved iron decreased towards those of the surrounding waters (Fig. 3a). After the final infusion (no. 4) levels of dissolved iron again initially decreased, but then remained at about 1 nM until the end of the site occupation. We note that Fe(II) —estimated to have a half-life of less than 1.0 h for oxidation to Fe(III) in these waters²⁵—was the predominant species on days 12/13. Based on our limited understanding of iron chemistry in the ocean²⁶, the high concentrations of Fe(II) and their persistence over several day/night cycles is best explained by a production mechanism (photoreduction) coupled with a maintenance mechanism (iron complexation), respectively. The existence of the latter mechanism is supported by a small data set (relative to that for Fe(II)) of Fe-binding ligand measurements; concentrations of such ligands remained at about 3.5 nM both inside and outside the patch until day 11, then increased to 8.5 nM inside the patch (P.C., unpublished data).

Bloom evolution

The first indication of an algal response to iron enrichment was a small but detectable increase in F_v/F_m after 24 h (Fig. 3b), with no change in the magnitude of algal stocks or production until days 3/4 (Fig. 3c and d). Subsequent sustained increases in both chlorophyll *a* and production rates in the patch were not reflected by trends in F_v/F_m , which increased immediately after each infusion, followed by a small decrease within 24 h (Fig. 3b). Although F_v/F_m increased by day 1, other physiological indices of the alleviation of algal iron stress—slower saturating rates of iron uptake (M.M., unpublished data), and decreases in diatom flavodoxin levels²¹ and in algal sinking rates²⁷ (Table 1)—were not evident until days 5–7. This time lag suggests that cells initially allocated resources to components with

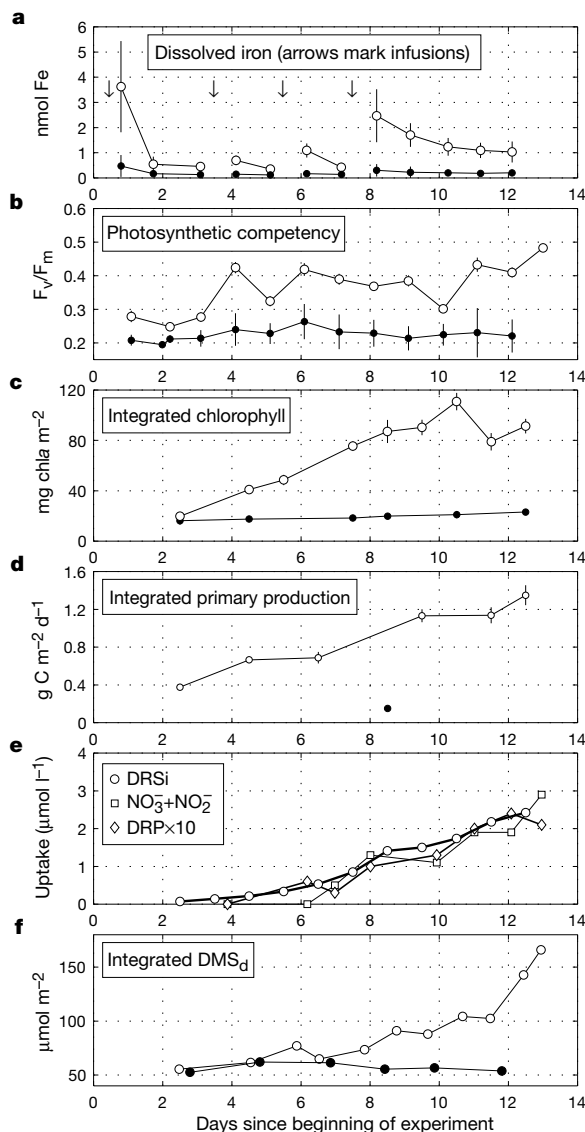


Figure 3 Time-series measurements made during SOIREE. Open and filled symbols show measurements made respectively inside and outside iron-fertilized waters. Underway data are expressed as the daily mean value for inside (defined as $>50\%$ of the peak SF_6 levels on that day) and outside the patch (defined as $<10 \text{ fmol l}^{-1}$). Error estimates are expressed as ± 1 standard deviation of the mean for underway data (a, b and e) and ± 1 standard error of the mean for the discrete data (c, d and f). The range of sample sizes during SOIREE is provided in parentheses: a, underway sampling of dissolved iron levels ($n = 5\text{--}17$ (in); $9\text{--}29$ (out)) in surface waters ($\sim 3 \text{ m}$) and the timing of the four iron infusions; b, underway sampling ($\sim 5 \text{ m}$) of community photosynthetic competency (F_v/F_m) ($n = 8\text{--}40$ (in); $6\text{--}84$ (out)); c, column-integrated chlorophyll *a* (six depths, $0\text{--}65 \text{ m}$, $n = 3$ pseudo-replicates); d, column-integrated primary production (six depths, $0\text{--}65 \text{ m}$, $n = 3$ pseudo-replicates); e, macronutrient uptake (silicate: from ^{32}Si uptake³³, $n = 2$ pseudo-replicates, error always $<10\%$ of the mean), nitrate and phosphate depletion (underway sampling ($\sim 5 \text{ m}$), $n = 12\text{--}18$ (in); $14\text{--}34$ (out)), the greatest spatial variability was $<15\%$ of the standard deviation of the mean), DRSi is dissolved reactive silica, DRP is dissolved reactive phosphorus; f, column-integrated DMS (six depths, $0\text{--}65 \text{ m}$, $n = 3$ pseudo-replicates, errors are all smaller than the symbol size). Nitrate and phosphate depletion was obtained by subtraction of levels within the patch from levels in the surrounding waters (which exhibited no significant change in during SOIREE).

an obligate requirement for iron²⁸, reflecting different response times of physiological and biochemical processes to iron enrichment. Inside the patch, net algal growth (see Methods) increased from $0.08 \pm 0.01 \text{ d}^{-1}$ early in SOIREE to $0.19 \pm 0.02 \text{ d}^{-1}$ later in the study, that is, lower than in on-deck studies¹⁰. These elevated growth rates resulted in a sixfold increase in chlorophyll *a* (Fig. 2) whereas algal carbon tripled. Thus, ratios of carbon to chlorophyll *a* that were approximately 90 just before SOIREE halved by day 13 within the patch. Surveys of surrounding waters indicated no significant changes in properties during SOIREE (Figs 2 and 3).

In the upper ocean, properties such as particle attenuation (the effect of particle density on transmission of light through sea water) were initially vertically homogeneous within the patch, but became heterogeneous over time, exhibiting a subsurface maximum by day 13 (Fig. 4). This transition was attributed to the presence of thermal features after day 5 (Fig. 1d). Settling of large cells from the upper 20 m is unlikely to explain this heterogeneity, as their sinking rates decreased over time. Alternatively, the trapping of cells in this upper layer probably resulted in marked photoinhibition (relative F_v/F_o , a proxy for photoinhibition²⁹ (see Methods), was 0.5 on days 7/8), which cells at depths in excess of 20 m did not exhibit (relative $F_v/F_o > 1.0$). Modelling and experimental studies on Antarctic phytoplankton indicate that inhibition of photosynthesis by ultraviolet radiation is relatively high during such transient events³⁰. Exposure to higher irradiances and elevated ultraviolet dosage may have depressed the F_v/F_m values of cells in surface waters, relative to those at depth (Fig. 4). Consequently, the magnitude of F_v/F_m from underway mapping may not be representative of cells at depth after day 5 (Fig. 3b).

Water column optics and the role of light availability

The development of the bloom resulted in reduced mean water-column light levels, reducing the photic depth I_0 (nominally 1% of surface photosynthetically active radiation (PAR)), from 84 m ($K_d = 0.055 \text{ m}^{-1}$, see Methods) to about 45 m ($K_d = 0.105 \text{ m}^{-1}$) over 13 d. Such self-shading may have altered algal iron requirements during SOIREE^{14,31}, and can impose an upper limit on chlorophyll *a* levels^{12,13}. An on-deck study (see Methods) investigated the effect of mixed-layer depth (see Fig. 1a), and thus mean irradiances, on iron-enriched phytoplankton. Although based on a limited number of samples, the '40 m' and '65 m' treatments both exhibited similar increases in chlorophyll *a* (Fig. 5) and macronutrient depletion (not shown) that were comparable to (in fact, twice) those measured *in situ*. In contrast, in the '100 m' treatment there was little increase in chlorophyll *a*: this suggests light limitation, or Fe/light co-limitation, of growth in waters with deep mixed layers.

Pelagic community response

The algal and heterotrophic communities responded differently to iron enrichment. Initial increases in chlorophyll *a* and cell abundance were due to pico-eukaryotes (Table 2), and abundances of autotrophic flagellate (mainly prymnesiophytes) increased between

days 2 and 8. However, after day 6, there was a floristic shift to large diatoms, which were dominated by *Fragilariopsis kerguelensis* ($30\text{--}50 \mu\text{m}$ cell length, $4.4 \times 10^4 \text{ cells l}^{-1}$ by day 12). The number of cells per chain for this species doubled to 14 by day 12 of SOIREE. Blooms of this species have been reported at the polar front in the Atlantic sector³². These floristic shifts probably mediated changes in the concentrations of climate-reactive gases in surface waters (see below). Moreover, iron supply to diatoms resulted in decreased Si:C uptake ratios³³, as observed previously in on-deck iron enrichments³⁴.

Between days 0 and 13 (Table 2), bacterial abundances changed little whereas production rates approximately tripled, suggesting significant rates of bacterivory. The timing of increases in bacterial production did not correspond to that of the initial iron enrichments, but followed increases in primary production (days 7–12). Ciliate abundances quadrupled by day 13 (Table 2), and microzooplankton herbivory was predominantly on cells smaller than $20 \mu\text{m}$, with little evidence of grazing on large diatoms. Copepods were the dominant large grazers, with krill and salps rare and absent, respectively, in net hauls. The clearance rates of the copepods changed little, and although chlorophyll *a* increased inside the patch, resulting in elevated copepod ingestion rates, these remained relatively low. The dominance of the bloom by *F. kerguelensis*, a highly silicified diatom that may be morphologically adapted to minimize grazing³⁵, was probably responsible for low rates of diatom herbivory.

Biogeochemical response

Iron-increased primary production resulted in significant uptake of macronutrients (Fig. 3e), and decreases in both partial pressure of CO_2 (p_{CO_2} ; Fig. 2) and dissolved inorganic carbon ($12\text{--}18 \mu\text{mol kg}^{-1}$)³³ from days 5/6 onwards. There was no significant change in dissolved nitrous oxide levels in the patch relative to the surrounding waters (data not shown). During SOIREE, large diatoms accounted for up to 75% of primary production (see Methods) and were probably responsible for most of the observed $35\text{--}40 \mu\text{atm}$ decrease in p_{CO_2} (ref. 33). Particulate dimethylsulphoniopropionate (DMSPp; the algal precursor of dimethyl sulphide, DMS) levels had increased after day 3, whereas dissolved DMS (DMS_d) concentrations initially increased on day 8 within the patch, and had tripled by day 13 (Fig. 3f). In contrast to the mainly diatom-mediated shifts in p_{CO_2} , other plankton are probably responsible for elevated DMS levels; increases in prymnesiophyte

Table 1 Particulate export and algal settling estimates during SOIREE

Time periods	Value in-patch/value outside-patch			
	Trap POC flux	Trap BSi flux	$^{234}\text{Th}:$ ^{238}U	Mean phytoplankton (>10 μm) sinking rates
Days 0–2	–	–	1.06 (± 0.14)	1.31
Days 7–9	1.18	0.94	1.04 (± 0.12)	0.70
Days 11–13	1.64	2.21	0.99 (± 0.14)	0.57

POC and BSi denote downward particulate organic carbon and biogenic silica, respectively. All data are expressed as relative ratios of the magnitude of export estimates (at 100 m depth) and algal sinking rates (measured on samples from 5 m depth) inside and outside the iron-fertilized patch over time periods when free-drifting sediment traps were deployed. $^{234}\text{Th}:$ ^{238}U represents the integrated (0–100 m) $^{234}\text{Th}:$ ^{238}U activity ratio, expressed here as a relative in-patch/outside-patch difference and the associated error (± 1 standard error of the mean). Errors cannot be calculated for trap fluxes and algal sinking rates as only one datum was obtained from outside-patch stations per deployment.

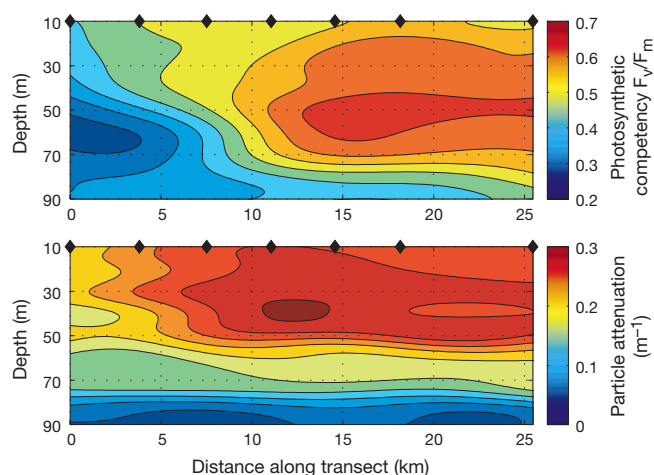


Figure 4 A vertical section (from the southeast to the northwest) through the iron-enriched patch on the night of 22 February. See SF₆ (day 12) in Fig. 2 for transect location in relation to iron-fertilized waters. **a**, Photosynthetic competency (F_v/F_m), and **b**, particle attenuation (background attenuation due to seawater has been subtracted). The filled diamonds denote CTD casts.

abundances, and subsequent decreases—owing to elevated herbivory by microzooplankton—were coincident with marked increases in DMSP (days 3–8), and in DMS (days 9–13), respectively (S.T., unpublished data). Laboratory studies have shown that the prymnesiophyte–microzooplankton pathway is particularly important in the production of DMSP and subsequent release of DMS³⁶. Thus, upper-ocean concentrations of climate-reactive gases may be controlled by different algal taxa during iron-mediated blooms.

Particle export was estimated using surface-tethered free-drifting sediment traps, and the deficit of total dissolved and particulate ²³⁴Th in the upper 100 m. There was no evidence of iron-increased particle export from the ²³⁴Th data (Table 1). The trap collections exhibited greater export of particulate organic carbon and biogenic silica inside the patch than outside (Table 1). However, large variations in fluxes of particulate organic carbon (75%) and biogenic silica (17%) were recorded during two deployments outside the patch (days 0–2 and 11–13), suggesting that the apparent increases inside the patch could result from the inherent variability of the trapping methodology³⁷. The decrease in algal sinking rates within iron-enriched waters, which has been observed in laboratory cultures when iron-depleted diatoms were iron-enriched²⁷, also supports the conclusion that there was no discernible increased export rates during the brief 13-d observation period.

Fate and longevity of the bloom

We now consider the physiological status of the algal community within the patch on day 13 of SOIREE. Algal F_v/F_m was sub-maximal in the upper 15 m (0.5; ref. 4), but maximal (0.65; ref. 4) at 50 m (Fig. 4), suggesting that cells were not resource-limited at depth. The high ambient nitrate and phosphate levels in surface waters were unlikely to limit growth, and there is indirect evidence from an on-deck experiment that cells were silicate-replete; the silicate uptake kinetics of the resident diatoms permitted silicate depletion to at least 4 μM , well below the ambient levels ($> 6 \mu\text{M}$) detected on day 13. Levels of dissolved iron on day 13 were ten times higher than in surrounding waters (Fig. 3a) yet the flavodoxin levels of diatoms (0–10 m) in iron-fertilized waters, which reached a minimum of about 50% of the initial value between days 5 and 8, increased after day 8 until the end of SOIREE.

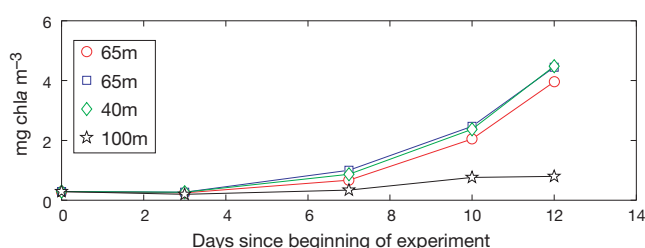


Figure 5 Changes in chlorophyll *a* levels with time during an on-deck iron/light perturbation experiment. Iron-enriched samples were incubated at a range of mean light levels simulating mixed-layer depths of 40 m (20% I_0), 65 m (10% I_0), two replicate carboys and 100 m (5% I_0) (see Methods). The standard error of the mean of three pseudo-replicates from each carboy was $<12\%$ in all cases.

The cellular accumulation of flavodoxin²¹ is particularly sensitive to algal iron stress, being detectable in cultured diatoms at 90% of maximum algal growth rates³⁸. Moreover, flavodoxin and F_v/F_m respond to iron limitation on different timescales, and represent different classes of photosynthetic response. Any iron-mediated down-regulation of photosynthetic capacity, reflected by a reduction in F_v/F_m , is likely to be delayed until all other compensatory mechanisms to reduce cellular iron requirements, such as flavodoxin accumulation, have been exploited³⁸. The presence of flavodoxin suggests that bioavailable iron levels were sub-optimal on day 13, at least in the dominant diatom species. These conditions may have been caused by changes in the partitioning of iron between the inorganic and organic pools. Rates of iron uptake from siderophore-bound iron, which are enhanced in iron-limited cells³⁹, were faster towards the end of SOIREE (M.T.M., unpublished data), suggesting that cells were iron-stressed. In spite of increased algal iron stress, primary production rates were high, and by inference the bloom remained productive after day 13 (ref. 40).

Levels of chlorophyll *a* increased by more than six times during SOIREE, yet there were no significant concomitant increases in the magnitude of loss processes, with small losses to the grazer and export routes (Tables 1 and 2). Thus, at the patch centre the accumulation of algal carbon (4.1 g C m^{-2} , see Methods) accounted for approximately 67% of iron-increased primary production (6.1 g C m^{-2} , see Fig. 3d). Application of these data to the 200-km² patch provides an upper estimate of carbon accumulation by day 13 (800 t algal C, see Methods) in response to the addition of about 1.7 t of iron. This is less than calculated by other methods³³, illustrating the uncertainties in making such estimates. On day 13, the 200-km² patch could potentially be distinguished from surrounding HNLC waters using SeaWiFS 1-km-resolution Ocean Color satellite data. On 21/22 March 1999, a SeaWiFS image showed a prominent ribbon-shaped feature (about 1,100 km²) with estimated chlorophyll *a* concentrations of up to 3 mg m^{-3} (ref. 40). Calculations of the physical and biological evolution of the patch after our departure, and its geographical proximity to the location of the SOIREE patch on 22 February, suggest that the provenance of this feature is the iron-fertilized bloom⁴⁰. This may alter the magnitude of biogeochemical signals such as p_{CO_2} draw-down within this feature, relative to the amount of iron added. The accumulation of algal carbon in the bloom by late March may be considerably larger than estimated on day 13 (ref. 40).

SOIREE and the iron hypothesis

The ‘ecumenical iron hypothesis’⁴¹ summarizes our understanding of the relative roles of iron supply, uptake, algal growth and community structure, and grazing in the functioning of HNLC regions. Confirmation of this hypothesis⁴¹ in Southern Ocean waters may be difficult, because light^{12,13}, iron/light co-limitation¹⁴, or iron and silicate availability (V. Franck, personal communication) may also exert control on phytoplankton processes. At the onset of SOIREE, owing to the presence of a relatively deep mixed layer and intermediate silicate concentrations, we could not discount the influence of such factors on algal growth at this site. However, indirect evidence from *in situ* and on-ship observations

Table 2 Response of the microbial and metazoan communities during SOIREE

Parameter	Inside-patch abundances and rates	Outside-patch abundances and rates
Heterotrophic bacterial abundance	$(4 \pm 0.1) \times 10^8 \text{ cells l}^{-1}$	$(3 \pm 0.1) \times 10^8 \text{ cells l}^{-1}$
Heterotrophic bacterial production	$(50 \pm 8) \text{ ng C l}^{-1} \text{ h}^{-1}$	$(20 \pm 2) \text{ ng C l}^{-1} \text{ h}^{-1}$
Heterotrophic ciliate abundance	$(13 \pm 4.6) \times 10^2 \text{ l}^{-1}$	$(3 \pm 0.5) \times 10^2 \text{ l}^{-1}$
Pico-eukaryote abundance	$(17 \pm 2.9) \times 10^6 \text{ l}^{-1}$	$(5 \pm 0.6) \times 10^6 \text{ l}^{-1}$
Mesozooplankton stocks (0–70 m)	$(2.0 \pm 0.59) \text{ g C m}^{-2}$	$(1.6 \pm 0.86) \text{ g C m}^{-2}$
Mesozooplankton herbivory* (0–65 m)	$(0.7 \pm 0.12) \% \text{ chlorophyll } a \text{ cleared d}^{-1}$	$(0.7 \pm 0.28) \% \text{ chlorophyll } a \text{ cleared d}^{-1}$

*Estimated using gut fluorescence⁵⁰. ‘Inside-patch’ denotes the highest daily mean (and standard error of the mean, based on at least three pseudo-replicates) of the abundances or rates observed in iron-fertilized waters during SOIREE. ‘Outside-patch’ denotes mean and standard error of abundances or rates ($n > 3$) for all days sampled. No significant changes in bacterial or meso- and micro-zooplankton abundances or activity were recorded in the surrounding waters. A comparison of day- and night-time mesozooplankton net hauls provided no evidence of arrested vertical migration within the patch relative to outside.

ruled out irradiance and silicate supply as significant factors during the 13-d study. During SOIREE, significant increases in algal F_v/F_m , growth, stocks and production rates provide unequivocal evidence of the role of iron in controlling the magnitude of phytoplankton processes. Moreover, iron supply—in conjunction with grazer dynamics—led to floristic shifts that resulted in a diatom-dominated bloom. These findings confirm the ecumenical iron hypothesis⁴¹ for this site in summer. Although iron-mediated increases in primary production and a floristic shift to large diatoms will tend to elevate export production rates⁴², this was not observed during our brief site occupation. Therefore, the second tenet of Martin's iron hypothesis¹—regarding iron supply and the subsequent enhancement of downward particulate organic carbon export—could not be confirmed.

Extrapolation to Southern Ocean waters

We will now put the SOIREE results into a wider context. Although Southern Ocean waters are characterized by strong meridional gradients¹⁶, the SOIREE site (65-m mixed layer; silicate, 10 μM) is broadly representative of more than 75% of polar waters in the Australian-Pacific sector of the Southern Ocean in summer (see Fig. 1a). Thus, our main finding—iron supply controls phytoplankton growth—may be cautiously applied to these waters. The ability to relate our findings, for summer, to other Southern Ocean sectors will require information on mixed-layer depth⁴³ (see Fig. 5) and on the relationship between ambient silicate levels and algal silicate uptake kinetics¹⁴.

Geochemical models of sustained iron enrichment of the Southern Ocean⁴⁴, in the absence of data, assume an algal growth season of 3–12 months. SOIREE was conducted in late summer, when irradiances⁴⁵ and silicate levels¹⁸ are less than those in December; mixed-layer depths are^{20,46} similar to those in December. Thus, iron supply will probably control algal growth, under ambient conditions, during summer at this site. Evidence of continued growth⁴⁰ during March—despite incident irradiances of 50% of December levels¹³—suggests the 'iron-limited' growth season may be up to 4 months before irradiances severely limit algal growth. Based on silicate uptake over 13 d (3 μM , Fig. 3e) and winter reserve levels (20 μM)¹⁸ at this site, silicate supply will probably influence algal growth rates under a simulated season of iron enrichment⁴⁴. Furthermore, iron-mediated shifts in carbon/silicate uptake stoichiometry during SOIREE are an important determinant of the magnitude of carbon sequestration predicted by a modelling study of iron enrichment of the Southern Ocean in the geological past³³.

Southern Ocean bloom dynamics

Iron-increased chlorophyll *a* concentrations (more than 2 mg m^{-3}) during SOIREE are similar to those observed for polar front blooms from both vessel¹¹ and remote-sensing⁴⁷ surveys. Furthermore, the 13-d timescale required to attain 2 mg chl a m^{-3} during SOIREE is comparable to that estimated for blooms at the polar front using remote sensing⁴⁷. There has been much debate about what factors limit algal biomass in the open Southern Ocean^{1,10,11}, but less about what sets an upper limit for chlorophyll *a* levels. Bio-optical models^{12,13} predict an upper limit for chlorophyll *a* accumulation (1 mg m^{-3}) in the deep mixed layers (deeper than 60 m) near the polar front. Yet there is evidence from remote-sensing archives⁴⁷ of higher chlorophyll *a* levels (1.5–4 mg m^{-3}) for these waters than predicted^{12,13}. The findings of SOIREE permit a re-evaluation of these modelling studies. Although they^{12,13} could not consider the effects of iron supply on algal processes, Nelson and Smith¹³ indicate the importance of the magnitude of loss processes in setting the critical depth, and the subsequent relationship between mixed layer and critical depths. Based on our findings, iron-enrichment by reducing the magnitude of loss processes—both directly (decreasing algal sinking rates) and indirectly (floristic shifts to taxa less

prone to grazing³⁵)—permits chlorophyll *a* accumulation of greater than 1 mg m^{-3} .

Remotely sensed observations near the polar front⁴⁷ provide information on the geographical extent and timescales of algal blooms in polar waters. However, little is known about what determines the longevity of blooms in open Southern Ocean waters. The presence of high chlorophyll *a* concentrations more than 40 d after the onset of SOIREE raises fundamental questions about the mechanisms for the sustenance and longevity of this bloom. It is possible that the relatively short length scale of artificial relative to natural fertilization events may result in artefacts, such as the entrainment of silicate-rich waters as the SOIREE patch spread⁴⁰, that may have influenced bloom longevity. Nevertheless, a comparison of calculated algal iron requirements and iron levels within the SOIREE patch over 40 d indicate that biomass accumulation was sustainable⁴⁰. The mechanisms by which sufficient Fe, via recycling or retention within surface waters and/or biota, was maintained to permit continued algal growth are unclear. However, the ability of resident cells to access this pool of organically bound iron only partially may have contributed to such longevity (M.T.M., unpublished data).

Future studies

The findings of SOIREE confirm the 'ecumenical iron hypothesis'⁴¹ at this site in summer, and provide insights into the role of iron supply on bloom dynamics in the Southern Ocean. However, despite direct monitoring of the SOIREE bloom for 13 d, and indirectly⁴⁰ for a further 30 d, we can only speculate as to the fate—whether via export, remineralization and/or subduction—of the accumulated algal carbon. Thus, although we are able to justify relating our main findings to other periods of summer at this location, and to other polar waters, we cannot extrapolate them to longer-term sequestration of carbon in the Southern Ocean. Owing to the exchange of iron-fertilized and HNLC waters⁴⁰, SOIREE may be likened to a laboratory-culture chemostat in which silicate was added continuously, large diatoms were 'washed out'—resulting in both decreases in cell abundances (see ref. 48) and self-shading^{12,13}—yet iron levels were always sufficient to permit net algal growth. It is not possible to determine whether the longevity of the bloom, and hence the potential delay in the onset of enhanced export of particulate organic carbon, was due to such an exchange of waters and/or to the maintenance of increased concentrations of dissolved iron. In order to examine the role of iron supply on the magnitude of downward fluxes of particulate organic carbon, the design of future studies will need to address these uncertainties. Furthermore, other approaches—such as modelling—will be required to constrain better the role of subduction in the eventual fate of iron-increased algal carbon in the Southern Ocean. □

Methods

Site selection and survey

Examination of climatological data sets (such as bathymetry, mixed-layer depth, buoy drift trajectories), and repeat hydrographic¹⁹ and XBT sections²⁰ suggested that 61°S, 140°E was a potential site. Satellite sea surface height and Ocean Colour observations indicated that in January 1999 this region had low eddy activity and low chlorophyll *a* levels. XBT sections (RV *Astrolab*) along ~142°E, obtained 30 d and 14 d before SOIREE, re-confirmed the geographical location of the main frontal boundaries. During our pre-release survey, chlorophyll *a*, F_v/F_m and macronutrients were sampled from the vessel's non-toxic seawater supply (intake 5 m subsurface) and analysed¹⁴ respectively by fluorometry (calibrated periodically with discrete chlorophyll *a* samples, corrected for quenching during daylight hours), fast-repetition-rate fluorometry, samples dark-adapted for 5 min, and using automated nutrient analysis. Underway samples for dissolved iron were obtained using a towed fish sampler, and were drawn from the vessel's non-toxic seawater system for $p\text{CO}_2$. Samples for the single-cell and community flavodoxin assay²¹ were pre-concentrated on board the vessel and analysed on land. There are several caveats associated with this assay¹⁴.

Initiation of the SF_6/Fe patch

The SF_6 solution (~0.32 mmol l^{-1}) was prepared on the outward transit leg by saturating

4,000-litre tanks of sea water with pure SF_6 (ref. 24). During infusion no. 1, 3,600 litres of sea water containing $\sim 165 \text{ g}$ (1.1 mol) SF_6 were mixed with $\sim 28,000$ litres of sea water containing $3,813 \text{ kg}$ $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (ref. 24) using a dosing unit to control the mixing and pumping rate of Fe/SF_6 into surface waters (at $\sim 15\text{-m}$ depth using a depressor). The release was co-ordinated within a lagrangian framework to account for surface water advection²⁴. A WOCE-type surface drifting buoy attached to a holey-sock drogue centred at 30 m depth and fitted with GPS and argos was used to mark the centre-point of the patch²⁴. The buoy position was updated every 10 min by UHF radio link²⁴. The release track was based upon an expanding hexagon around the centre-point with track spacings at 600-m intervals, and a ship speed of 2–3 knots. SF_6 concentrations were analysed during daily underway mapping of the patch (5 m subsurface, vessel's non-toxic seawater system) and from vertical profiles taken from daily hydro-casts⁴⁹. During infusion no. 1, the iron sulphate added to the upper ocean resulted in a theoretical addition of 3.8 nM Fe in a 50 km^2 patch. For infusions nos 2 and 3, $1,550 \text{ kg}$ $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (in $12,000$ litres of acidified seawater) was added each time (theoretical addition 2.6 nM Fe), and during the final infusion (no. 4) $1,750 \text{ kg}$ $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (in $8,000$ litres of acidified sea water) was added (theoretical addition 2.5 nM Fe).

Hydrographic sampling

The upper 150 m was sampled vertically (6–8 depths) each day using 12 acid-cleaned 10-litre bottles on a CTD rosette system at the patch centre (determined by SF_6 levels) and in the surrounding waters. Real-time vertical profiling of temperature, salinity, transmissivity, F_0/F_m , chlorophyll *a* fluorescence and underwater irradiance (PAR, 380–700 nm)¹⁴ was performed. The attenuation coefficient (K_d) for PAR is not constant with depth in the upper ocean and was obtained by curve-fitting. F_0/F_m , an algebraic transform of F_0/F_m , was obtained from vertical fast-repetition-rate fluorometry casts. Discrete water samples were analysed for chlorophyll *a* (size-fractionated; >20 , 5–20, 2–5 and $0.2\text{--}2 \mu\text{m}$)¹⁴, heterotrophic bacterial abundance⁵⁰, microzooplankton abundance⁵⁰, and phytoplankton abundance (see below)⁵⁰. Additional samples were incubated on deck to measure rates of primary production (^{14}C , 24-h incubation, simulated *in situ*, size-fractionated as for chlorophyll *a*)¹⁴, bacterial production⁵⁰, and microzooplankton grazing⁵⁰. Estimates of cell size for each algal group in conjunction with size-fractionated chlorophyll *a* production were used to assess the contribution of algal taxa to community biomass or production. Mesozooplankton abundance was assessed from vertical hauls⁵⁰ from which additional animals were obtained and used, after 24 h acclimation, in herbivory experiments⁵⁰. The Th:U activity ratio of particles in the upper ocean were collected using a subsmersible pumping system.

On-deck perturbation experiments

Water for an iron/light experiment was collected on 10 February from the patch centre (dissolved iron, 2 nM) using 30-litre clean Go-Flo samplers⁵¹. Sea water was added to clean⁵¹ 25-litre polycarbonate carboys, within a clean air (class 100) environment, which were then shaded with neutral density screening to simulate the calculated mean light levels (see caveats⁵²) in a 40-, 65- or 100-m mixed layer. Carboys were incubated in Perspex vessels (at $\sim 2^\circ\text{C}$) for 12 d, and periodically subsampled (class 100 environment). Sub-samples were analysed for chlorophyll *a* and macronutrient levels.

Algal carbon accumulation, C:chlorophyll *a*, and growth rates

Algal cells from four depths (0–65 m) at the patch centre were identified, enumerated and sized using light (diatoms/dinoflagellates) or epifluorescence (autotrophic flagellates) microscopy, and flow cytometry (pico-eukaryotes, verified with epifluorescence microscopy), then converted to biovolume and cellular carbon using size and taxon-specific factors⁵³. Algal carbon was integrated over the 65-m mixed layer. The accumulation of algal carbon during SOIREE was obtained by subtracting the column-integral on day 13 from that at the onset. Extrapolation of column-integrated algal carbon from the patch centre to the 200-km^2 patch yielded an upper limit of algal C accumulation during SOIREE (800 t C). However, if algal carbon was distributed throughout the patch in the same manner as SF_6 , this estimate would be $\sim 400 \text{ t C}$. Algal C and primary production data were used to estimate algal net growth rates⁵⁴ by comparing the cumulative iron-mediated increase in column-integrated production over SOIREE to the column-integrated accumulation of algal C. Column-integrated chlorophyll *a* and algal carbon were used to provide C:chlorophyll *a* ratios.

Received 6 January; accepted 4 September 2000.

- Martin, J. H. Glacial-interglacial CO_2 change: The iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–206 (1998).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
- Behrenfeld, M. J., Bale, A. J., Kolber, Z. S., Aiken, J. & Falkowski, P. G. Confirmation of iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature* **383**, 508–510 (1996).
- Cooper, D. J., Watson, A. J. & Nightingale, P. D. Large decrease in ocean-surface CO_2 fugacity in response to *in situ* iron fertilization. *Nature* **383**, 511–513 (1996).
- Sarmiento, J. L. & Orr, J. C. Three-dimensional simulations of the impact of Southern Ocean nutrient depletion on atmospheric CO_2 and ocean chemistry. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
- Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J. & Manabe, S. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding termination II and their implications for the causes of glacial interglacial CO_2 changes. *Paleoceanography* **13**, 352–364 (1998).
- Kumar, N. *et al.* Increased biological productivity and export production in the glacial Southern Ocean. *Nature* **378**, 675–680 (1995).

- de Baar, H. J. W. & Boyd, P. W. in *The Dynamic Ocean Carbon Cycle: A Midterm Synthesis of the Joint Global Ocean Flux Study* Ch. 4 (eds Hanson, R. B., Ducklow, H. W. & Field, J. G.) 61–140 (International Geosphere Biosphere Programme Book Series, Cambridge Univ. Press, Cambridge, 1999).
- de Baar, H. J. W. *et al.* Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean. *Nature* **373**, 412–415 (1995).
- Mitchell, B. G., Brody, E. A., Holm-Hansen, O., McClain, C. & Bishop, J. Light limitation of phytoplankton biomass and macronutrient utilization in the Southern Ocean. *Limnol. Oceanogr.* **36**, 1662–1677 (1991).
- Nelson, D. M. & Smith, W. O. Jr Sverdrup revisited: Critical depths, maximum chlorophyll levels, and the control of Southern Ocean productivity by the irradiance-mixing regime. *Limnol. Oceanogr.* **36**, 1650–1661 (1991).
- Boyd, P. W., LaRoche, J., Gall, M., Frew, R. & McKay, R. M. L. The role of iron, light and silicate in controlling algal biomass in sub-Antarctic waters SE of New Zealand. *J. Geophys. Res.* **104**, 13395–13408 (1999).
- Banse, K. Iron availability, nitrate uptake and exportable new production in the subarctic Pacific. *J. Geophys. Res.* **96**, 741–748 (1991).
- Orsi, A. H., Whitworth, T. W. & Nowlin, W. D. Jr On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Res.* **1** **42**, 641–673 (1995).
- Nowlin, W. D. Jr & Klinck, J. M. The physics of the Antarctic Circumpolar Current. *Rev. Geophys. Space Phys.* **24**, 469–491 (1986).
- Zentara, S. J. & Kamykowski, D. Geographic variation in the relationship between silicic acid and nitrate in the South Pacific Ocean. *Deep-Sea Res.* **A** **28**, 455–465 (1981).
- Rintoul, S. R. & Bullister, J. L. A late winter hydrographic section from Tasmania to Antarctica. *Deep-Sea Res.* **1** **46**, 1417–1454 (1999).
- Rintoul, S. R., Donguy, J. R. & Roemmich, D. H. Seasonal evolution of upper ocean thermal structure between Tasmania and Antarctica. *Deep-Sea Res.* **1** **44**, 1185–1202 (1997).
- LaRoche, J., Boyd, P. W., McKay, R. M. L. & Geider, R. J. Flavodoxin as an *in situ* marker for iron stress in phytoplankton. *Nature* **382**, 802–805 (1996).
- Watson, A. J., Liss, P. S. & Duce, R. A. Design of a small-scale iron fertilisation experiment. *Limnol. Oceanogr.* **36**, 1960–1965 (1991).
- Martin, J. H. *et al.* Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**, 123–129 (1994).
- Law, C. S., Watson, A. J., Liddicoat, M. I. & Stanton, T. Sulphur hexafluoride as a tracer of biogeochemical and physical processes in an open-ocean iron fertilisation experiment. *Deep-Sea Res.* **II** **45**, 977–994 (1998).
- Millero, F. J., Sotolongo, S. & Izaguirre, M. The oxidation kinetics of Fe(II) in seawater. *Geochim. Cosmochim. Acta* **51**, 793–801 (1987).
- Geider, R. J. Biological oceanography: complex lessons of iron uptake. *Nature* **400**, 815–816 (1999).
- Muggli, D. L., Lecourt, M. & Harrison, P. J. Effects of iron and nitrogen source on the sinking rate, physiology and metal composition of an oceanic diatom from the subarctic Pacific. *Mar. Ecol. Prog. Ser.* **132**, 215–227 (1996).
- Greene, R. M., Geider, R. J., Kolber, Z. & Falkowski, P. G. Iron-induced changes in light-harvesting and photochemical energy conversion processes in eukaryotic marine algae. *Plant Physiol.* **100**, 565–575 (1992).
- Behrenfeld, M. J., Prasil, O., Kolber, Z. S., Babin, M. & Falkowski, P. G. Compensatory changes in photosystem II electron turnover rates protect photosynthesis from photoinhibition. *Photosynth. Res.* **58**, 259–268 (1999).
- Neale, P. J., Davis, R. F. & Cullen, J. J. Interactive effects of ozone depletion and vertical mixing on photosynthesis of Antarctic phytoplankton. *Nature* **392**, 585–589 (1998).
- Sunda, W. G. & Huntsman, S. A. Interrelated influence of iron, light and cell size on marine phytoplankton growth. *Nature* **390**, 389–392 (1997).
- Laubscher, R. K., Perissinotto, R. & McQuaid, C. D. Phytoplankton production and biomass at frontal zones in the Atlantic Sector of the Southern Ocean. *Polar Biol.* **13**, 471–481 (1993).
- Watson, A. J., Bakker, D. C. E., Ridgwell, A. J., Boyd, P. W. & Law, C. S. Effect of iron supply on Southern Ocean CO_2 uptake and implications for glacial atmospheric CO_2 . *Nature* **407**, 730–733 (2000).
- Hutchins, D. A. & Bruland, K. W. Iron limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**, 561–564 (1998).
- Verity, P. & Smetacek, V. Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Mar. Ecol. Prog. Ser.* **130**, 277–293 (1996).
- Wolfe, G. V., Steinke, M. & Kirst, G. O. Grazing-activated chemical defence in a unicellular marine alga. *Nature* **387**, 894–897 (1997).
- Buesseler, K. O. Do upper-ocean sediment traps provide an accurate record of particle flux? *Nature* **353**, 420–423 (1991).
- McKay, R. M. L., Geider, R. J. & LaRoche, J. Physiological and biochemical response of the photosynthetic apparatus of two marine diatoms to Fe stress. *Plant Physiol.* **114**, 615–622 (1997).
- Maldonado, M. T. & Price, N. M. Utilization of iron bound to strong organic ligands by phytoplankton communities in the subarctic Pacific Ocean. *Deep-Sea Res.* **II** **46**, 2447–2474 (1999).
- Abraham, E. R. *et al.* Importance of stirring in the development of an iron-fertilized phytoplankton bloom. *Nature* **407**, 727–730 (2000).
- Morel, F. M. M., Reuter, J. G. & Price, N. M. Iron nutrition of phytoplankton and its possible importance in the ecology of ocean regions with high nutrient and low biomass. *Oceanography* **4**, 56–61 (1991).
- Boyd, P. W. & Newton, P. P. Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces. *Deep-Sea Res.* **1** **46**, 63–91 (1999).
- Obata, A., J. Ishizaka & Endoh, M. Global verification of critical depth theory for phytoplankton blooms with climatological *in situ* temperature and satellite ocean color data. *J. Geophys. Res.* **101**, 20657–20667 (1998).
- Peng, T.-H. & Broecker, W. S. Factors limiting the reduction of atmospheric CO_2 by iron fertilisation. *Limnol. Oceanogr.* **36**, 1919–1927 (1991).
- Bishop, J. K. B. & Rossow, W. R. Spatial and temporal variability of global surface solar irradiance. *J. Geophys. Res.* **96**, 16839–16858 (1991).
- Trenberth, K. E., Large, W. G. & Olson, J. G. The mean annual cycle in global ocean wind stress. *J. Phys. Oceanogr.* **20**, 1742–1760 (1990).

47. Moore, J. K. *et al.* SeaWiFS satellite ocean color data from the Southern Ocean. *Geophys. Res. Lett.* **26**, 1465–1468 (1999).
48. Jackson, G. A. A model of the formation of marine algal flocs by physical coagulation processes. *Deep-Sea Res.* **37**, 1197–1211 (1990).
49. Law, C. S., Watson, A. J. & Liddicoat, M. I. Automated vacuum analysis of sulphur hexafluoride in seawater; derivation of the atmospheric trend (1970–1993) and potential as a transient tracer. *Mar. Chem.* **48**, 57–69 (1994).
50. Bradford-Grieve, J. M. *et al.* Pelagic ecosystem structure and functioning in the Subtropical Front region east of New Zealand in austral winter and spring 1993. *J. Plank. Res.* **21**, 405–428 (1999).
51. Fitzwater, S. E., Knauer, G. A. & Martin, J. H. Metal contamination and its effect on primary production methods. *Limnol. Oceanogr.* **27**, 544–551 (1982).
52. Maldonado, M., Boyd, P. W., Price N. M. & Harrison, P. J. Co-limitation of phytoplankton by light and Fe during winter in the NE subarctic Pacific Ocean. *Deep-Sea Res. II* **46**, 2475–2486 (1999).
53. Montagnes, D. J. S., Berges, J. A., Harrison, P. J. & Taylor, F. J. R. Estimating carbon, nitrogen, protein and chlorophyll *a* from volume in marine phytoplankton. *Limnol. Oceanogr.* **39**, 1044–1060 (1994).
54. Popp, B. N., Parekh, P., Tilbrook, B., Bidigare, R. R. & Laws, E. A. Organic carbon $\delta^{13}\text{C}$ variations in sedimentary rocks as chemostratigraphic and paleoenvironmental tools. *Palaeoclimatol. Palaeoecol.* **132**, 119–132 (1997).

Acknowledgements

We thank the officers and crew of the RV *Tangaroa*, and the management and staff of the NIWA Vessel Management Company. We also acknowledge the provision of physical oceanographic data by J. Church, M. Morris, V. Strass, RV *Astrolab*, and J. Barth and T. Cowles. We also thank M. Walkington (CTD calibration), J. Aiken (instrument loan), P. Nightingale (argos reception for tracking buoys), S. Groom (near-time high resolution Ocean Color satellite images), NASA (SeaWiFS images), and the companies CEFIC (UK) and BHP (Australia) for their support. We acknowledge two grants from the UK NERC to S.T. (Advanced Fellowship Award), and to A.J.W. and C.S.L. (SOIRE). K.O.B. and M.C. were funded by the US NSF. M.T.M. was funded by the NSERC Canada and the Center for Environmental Bioinorganic Chemistry, Princeton, USA. P.C., A.J.W. and D.C.E.B. received a grant within CARUSO from the European Union. New Zealand scientists were funded by the Government PGSF fund for Antarctic Research. J. Cullen, K. Hunter and C. Hurd provided comments and advice that improved this manuscript.

Correspondence and requests for materials should be addressed to P.W.B. (e-mail: pboyd@alkali.otago.ac.nz).

Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA

Galina Obmolova*, Changill Ban†‡, Peggy Hsieh* & Wei Yang†

* Genetics and Biochemistry Branch and † Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

DNA mismatch repair is critical for increasing replication fidelity in organisms ranging from bacteria to humans. MutS protein, a member of the ABC ATPase superfamily, recognizes mispaired and unpaired bases in duplex DNA and initiates mismatch repair. Mutations in human MutS genes cause a predisposition to hereditary nonpolyposis colorectal cancer as well as sporadic tumours. Here we report the crystal structures of a MutS protein and a complex of MutS with a heteroduplex DNA containing an unpaired base. The structures reveal the general architecture of members of the MutS family, an induced-fit mechanism of recognition between four domains of a MutS dimer and a heteroduplex kinked at the mismatch, a composite ATPase active site composed of residues from both MutS subunits, and a transmitter region connecting the mismatch-binding and ATPase domains. The crystal structures also provide a molecular framework for understanding hereditary nonpolyposis colorectal cancer mutations and for postulating testable roles of MutS.

DNA mismatch repair (MMR) primarily corrects mispaired or unpaired bases caused by DNA polymerase, thus increasing the overall fidelity of DNA replication by 100- to 1,000-fold^{1,2}. Cells harbouring mutations in MMR genes are characterized by greatly elevated rates of spontaneous mutation and instability of microsatellite repeats¹. Inactivation of MMR genes by mutation or epigenetic processes predisposes humans to hereditary nonpolyposis colorectal cancer (HNPCC) and sporadic tumours^{3–5}. In *Escherichia coli*, three MMR proteins, MutS, MutL and MutH, initiate a strand-specific mismatch repair process that takes advantage of the transiently unmethylated state of the newly synthesized daughter strand¹. MutS recognizes mispaired or 1–4 unpaired bases and, in an ATP-dependent fashion, activates the endonuclease MutH with the aid of another ATPase, MutL^{1,6}. Strand-specificity of mismatch repair resides in MutH, which recognizes a hemimethylated GATC sequence and cleaves only the unmethylated daughter strand. The GATC site can be over 1,000 base pairs (bp) away and on either side of the mismatch. Both MutS and MutL are also essential in the subsequent excision and resynthesis of the daughter strand.

Homologues of *E. coli* MutS have been found in nearly all organisms^{1,7}. Prokaryotic MutS proteins are encoded by a single gene and form homodimers^{8,9}. Eukaryotic genomes contain multiple *mutS* homologue (*msh*) genes. Except for mitochondrial MSH1, eukaryotic MutS proteins are heterodimeric. For instance, MSH2 pairs with MSH6 or MSH3 to form MutS α and MutS β ^{7,10}. MutS and its homologues can be divided into two lineages¹¹. The majority of prokaryotic and eukaryotic MutS proteins belong to the lineage that participates in MMR, prevents homologous DNA recombination between heterologous sequences, and mediates cell death induced by DNA damaging agents^{12,13}. The other lineage, including MSH4–MSH5, does not function in MMR, but promotes DNA pairing and crossing over during meiosis¹³.

All members of the MutS family possess a conserved ATPase activity^{1,7}. Both mismatch recognition and the ATPase activities of MutS are essential for MMR^{14–16} even though each activity is

independently detectable^{17,18}. However, it is unclear how MutS recognizes a broad range of mismatches embedded in random DNA sequence, how the MutS ATPase activity is involved, and how these two essential functions of MutS are related. We report here the crystal structures of *Thermus aquaticus* (TAQ) MutS and its complex with a heteroduplex DNA at 3.2 Å and 2.2 Å resolution, respectively, and discuss new insights into the function of MutS in MMR.

General features of MutS and a MutS–DNA complex

TAQ MutS, a thermostable homologue of *E. coli* MutS, comprises 811 amino-acid residues¹⁹. We have co-crystallized a nearly full-length TAQ MutS protein (residues 1–768) and a 21-bp heteroduplex containing a single, unpaired thymidine (T). The truncated TAQ MutS retains the dimerization, mismatch-recognition and ATPase activities of the full-length protein; a similarly truncated *E. coli* MutS is proficient in mismatch repair *in vivo* (P.H. *et al.*, unpublished data). The crystal contains one MutS dimer and one heteroduplex DNA in each asymmetric unit and diffracts X-rays to 2.0 Å resolution (see Table 1 in Supplementary Information). The co-crystal structure was phased by multiwavelength anomalous diffraction with selenylmethionyl substitution²⁰. The final refined model contains all nucleotides and residues 1–765 of MutS with the exception of 6 disordered residues, 629–634 (see Table 1 in Supplementary Information). The two subunits of MutS are related by a molecular pseudo-two-fold axis and form an oval disk (~125 × 90 × 70 Å; Fig. 1). Two adjacent large channels with dimensions of ~30 × 20 Å and ~40 × 20 Å penetrate the disk-like protein structure, and the latter is occupied by the heteroduplex DNA. The DNA is kinked sharply towards the major groove by ~60° at the unpaired T. Only one protein subunit interacts with the unpaired T, thereby breaking the molecular two-fold symmetry of the homodimer.

We crystallized a larger TAQ MutS fragment (residues 1–782) in the absence of DNA. The crystal structure was determined by molecular replacement and partially refined to 3.2 Å (see Fig. 1 in Supplementary Information). The protein-only structure is much more mobile than the MutS–DNA complex as indicated by much higher overall *B* values (see Table 1 in Supplementary Information).

‡ Present address: Hauptman-Woodward Medical Institute, Buffalo, New York, USA.

MutS protein

A MutS subunit consists of five structural domains arranged in the shape of a comma (Fig. 2a). Each of the five domains is structurally identical between the two subunits, even though the relative orientations of domains I, IV and V differ (Fig. 2b). Domains I (residues 1–118), II (residues 132–245), IV (residues 406–513) and V (residues 543–765) are of mixed α/β -fold, and domain III (residues 247–385 and 514–540) is entirely α -helical. Each of the five domains resembles at least one previously determined structure (see Fig. 2 in Supplementary Information). For example, domain II resembles RNase H and RNase H-like domains found in Hsc70, HIV integrase and DNA polymerase²¹; domain V, which contains the Walker A motif, is structurally similar to the ATPase domain of ABC transporters^{22,23}.

The structure of MutS is extended and contains multiple hydrophobic cores. Domain III is central to the MutS structure; it is

directly connected to domains II, IV and V by peptide bonds and shares the most extensive interdomain contacts with domain V. Domain IV is elongated, encompassing a pair of antiparallel α -helices over 40 Å in length with a small β -sheet at the distal end. The globular domain I and domain IV are involved in DNA binding. Each of the two domains is independently folded and appended to the core of MutS by flexible peptide linkages and limited interactions (Fig. 2), which indicate the possibility of domain rearrangement. The potential mobility of these domains is confirmed in the crystal structure of MutS alone. Although domains II, III and V retain similar structures in the presence or absence of DNA, domains I and IV are too mobile to be discerned in the absence of DNA (Fig. 3a).

MutS forms a stable dimer due to the extensive interactions between the ATPase domains, which buries more than 2,400 Å² of molecular surface (Fig. 3a, b). The presence of DNA brings domains

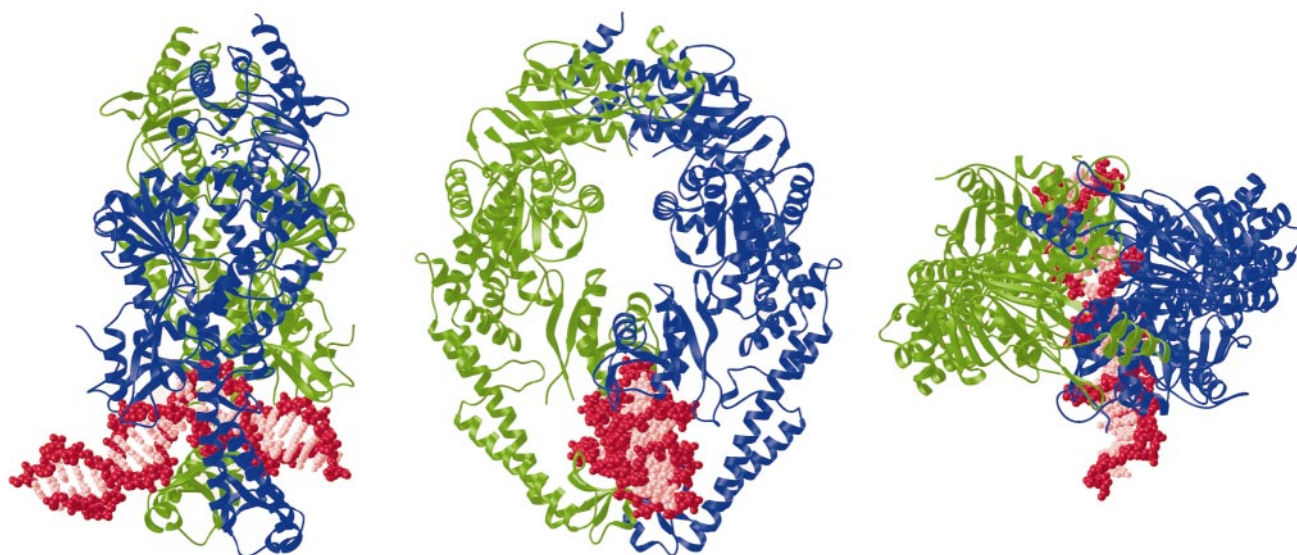


Figure 1 Crystal structure of the TAQ MutS–DNA complex. The two protein subunits are represented by ribbon diagrams in blue (A) and green (B). The DNA is shown in a space-

filling model, in which the backbone atoms are red and bases are pink. Three orthogonal views from the side, front and top are shown.

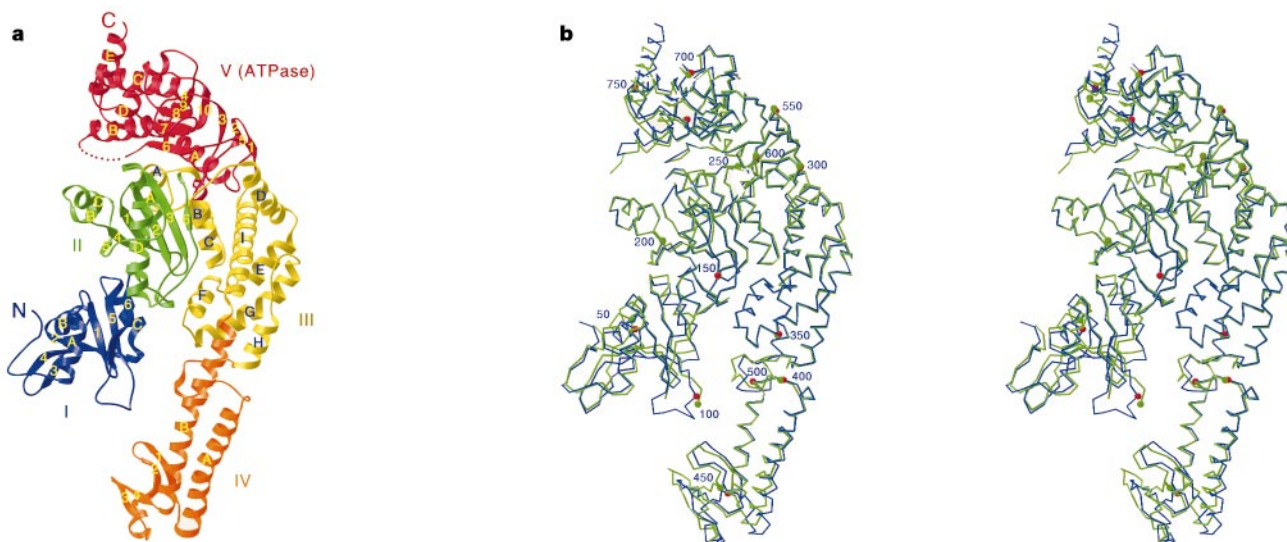


Figure 2 Crystal structure of a TAQ MutS subunit. **a**, Ribbon diagram of subunit A. The five domains are coloured blue, green, yellow, orange and red from the N to the C terminus. Secondary structures are named in alphabetic or numeric order for α -helices and β -strands, respectively, following their order in the primary sequence in each domain. **b**, A

stereo view of the superposition of C α traces of the two subunits. Subunit A is shown in blue and subunit B in green. Every fiftieth C α of each subunit is marked by a red (subunit A) or a green (subunit B) sphere.

I and IV into the dimer interface, which adds $\sim 600 \text{ \AA}^2$ to the buried surface and forms the lower channel that encircles the DNA (Fig. 3b). DNA binding also induces an outward rotation of the two MutS subunits (see Fig. 3 in Supplementary Information). The arrangement of the domains I and IV is reminiscent of the structures of DNA gyrase and topoisomerase II (ref. 24), which also contain globular DNA-binding domains linked by a pair of long anti-parallel helices and form a large channel for DNA passage. However, the two families of proteins share no further structural similarity.

At the dimer interface formed between domains V, a pair of anti-parallel helices (V α D and V α E) extends from the carboxy terminus of one subunit and interacts with a pair of parallel helices (V α B and V α C) from the other subunit (Fig. 3c). The interactions between these helices are mostly hydrophobic except for two hydrogen bonds involving the side chains of D672, S736 and the main chain of V739 (Fig. 3d), and they account for $\sim 70\%$ of the total dimer interface. Although V α D and V α E were thought to be involved in DNA binding by virtue of a conserved helix–turn–helix motif, mutational studies indicated that these two helices are involved in dimerization and ATPase activity¹⁵, a finding confirmed here. Removal of the last five residues of V α E abolishes not only dimerization of TAQ MutS, but also the ATPase and mismatch

binding activities (P.H. *et al.*, unpublished data). As V α E is far too distant to participate directly in DNA and ATP binding (70 Å and 20 Å, respectively), dimerization must be required for both activities.

Mismatch DNA recognition by MutS

The unpaired T in the MutS co-crystal structure is partially stacked in the DNA duplex (Fig. 4), in agreement with the observation that a flipped out nucleotide is not recognized by MutS²⁵. The base is displaced towards the minor groove by $\sim 3 \text{ \AA}$ and tilts towards the 5' base pair so that its O4 oxygen is hydrogen bonded to the N2 of the 5' G. This conformation of the unpaired T is stabilized by the interactions with MutS, which include an aromatic-ring stack with Phe 39 on the 3' side and a hydrogen bond with the side chain of Glu 41 (Fig. 4c). Phe 39 was postulated to be adjacent to the unpaired T based on ultraviolet crosslinking experiments²⁶. In the crystal structure, Phe 39 approaches the DNA from the minor-groove side, stacks onto the unpaired T, and contacts the base 3' to the T with the tip of its phenyl ring (Fig. 4c). As a result, the DNA is sharply kinked towards the major groove, and the minor groove is as wide as the adjacent major groove (Fig. 4a, b). All base pairs on either side of the kink assume a canonical B-form conformation except for a slight buckle at the base pair 5' to the unpaired T.

The kinked heteroduplex DNA is stabilized by extensive interactions with four MutS domains. Domains I and IV of the two subunits crisscross each other and surround 5 and 8 bp on each side of the kink, burying a total of $\sim 3,300 \text{ \AA}^2$ of molecular surface (Fig. 4a, b). Domains I and IV share a similar folding topology, two pairs of β -hairpins (β 1, β 2 and β 3, β 4) linked by a helical segment form an anti-parallel β -sheet (Fig. 4d). Similar anti-parallel β -sheets have been observed (see Fig. 2 in Supplementary Information), but this is the first example in which these β -sheets are found binding to DNA and interacting with either the major or the minor groove (Fig. 4). The protein–DNA interactions, except at the unpaired T, involve only the phosphate–sugar backbone of the DNA and thus impose no direct sequence specificity on mismatch recognition (Fig. 4e).

The interactions between MutS and DNA are asymmetric. Domain I of subunit A, which donates the intercalating Phe 39, interacts with nine nucleotides by wedging the β -sheet between phosphate backbones in the widened minor groove and anchoring a third β -hairpin (I β 5 and I β 6) on the outside (Fig. 4). The amino terminus of I α A also contributes to DNA binding (Fig. 4e). Many of these residues contacting phosphate–sugar backbones are conserved (Figs 4e, 5). Domain IV of subunit B, situated on the opposite side of DNA from domain I of subunit A, uses the two β -turns on the opposite ends of the β -sheet to contact the DNA backbones across the major groove, spanning at least 8 bp (Fig. 4). In contrast, domain IV of subunit A and domain I of subunit B, whose Phe 39 is not engaged in DNA binding and I β 5–I β 6 turn is partially disordered (Fig. 4a), contact only two or three adjacent nucleotides on a single strand (Fig. 4e). The asymmetric protein–DNA interactions surrounding the mismatch site and the sharp kink of DNA towards the major groove are in good agreement with previous DNase I and chemical footprinting results^{17,27}. The requirement for four domains from two protein subunits to bind DNA explains the dependence of mismatch recognition on protein dimerization¹⁵ (P.H. *et al.*, unpublished data).

The observed flexibility of MutS provides a structural basis for its broad substrate specificity. First, the mobility of the DNA-binding domains evident in the structure of MutS alone (Fig. 3a) provides an entry for DNA to the mismatch-binding channel. Second, the adaptability indicated by the asymmetric use of identical domains in DNA binding enables MutS to accommodate a broad range of heteroduplex DNAs, while maximizing the protein–DNA interface. MutS has been reported to bind DNAs containing mispaired or unpaired bases, 8-oxoG, O⁶meG or cisplatin adducts⁷, which are

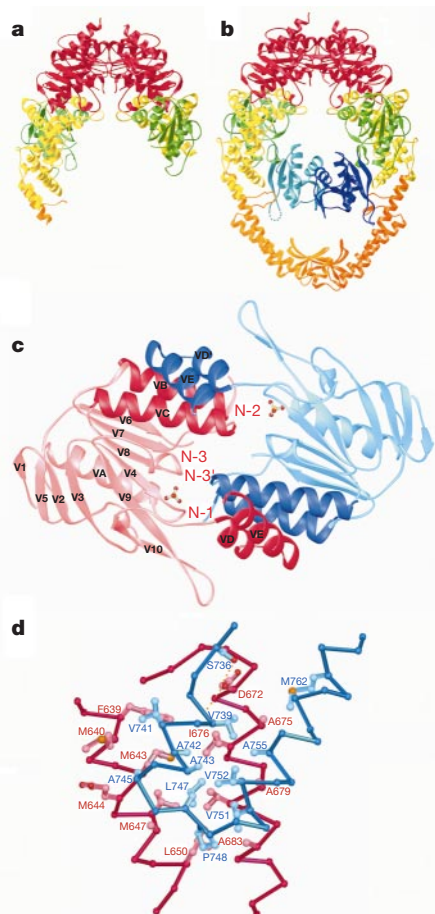


Figure 3 Structures of the MutS dimer. **a, b**, Ribbon diagram of the MutS dimer in the absence (**a**) or the presence (**b**) of the heteroduplex DNA. Each domain is shown in the same colours as in Fig. 2a. Domains I and IV of Subunit B are drawn in lighter colour than those of subunit A. **c**, Ribbon diagram of the dimer interface at domain V. Domain V of subunit A and B are coloured in blue and red, respectively. The helices that dominate the dimer interface are in a darker shade. The sulphate ions are shown as ball-and-stick. The four nucleotide-binding motifs in the red subunit are labelled. **d**, One of the four-helix bundles at the dimer interface. The mainchains are represented by the C α traces. The sidechains that interact between the two subunits are labelled and shown as ball-and-stick. Two hydrogen bonds are represented by yellow dashed lines.

likely to have different local structures. In support of this assessment, the replacement of the unpaired T with a T·T mismatch in the MutS–DNA complex results in a new crystal lattice (G.O. *et al.*, unpublished data).

Kinked DNA is likely to be a general feature in MutS–DNA complexes because a straight DNA would clash severely with one domain I and reduce the protein–DNA interface (Fig. 4a). Although mismatch-containing DNAs appear to be straight in the absence of protein, base stacking is less stable at a mismatch site and more susceptible to deformation²⁸. DNA modified by cisplatin is intrinsically kinked towards the major groove²⁹ and recognized by human MutS α ⁷. MutS proteins also bind nonspecifically to homoduplex DNA^{14,16}, which evidently can be bent or sharply kinked upon protein binding³⁰. A mutation of Phe 39 to Ala in TAQ MutS, or equivalent mutations in *E. coli* MutS and MSH6 of yeast MutS α , abolishes MutS binding to both hetero- and homoduplex DNA^{31,32} (A. Yamamoto and P.H., unpublished data) indicating that they may share a similar binding site in MutS.

The union of MutS and a heteroduplex DNA is thus a result of induced fit of both molecules that involves kinking of DNA and rearrangement of protein domains. The preferred binding of MutS to a heteroduplex DNA is most probably due to a reduction in the energetic cost of kinking DNA in the presence of a mismatch or a modified base²⁸. A propensity for stacking mismatched bases within

DNA helix rather than flipping them out may further explain the different binding efficiency of MutS for a T·G versus T·T mismatch and dependence on the surrounding base content, that is, AT- versus GC-rich sequence²⁵.

A composite ATPase active site

Four nucleotide-binding motifs (N-1, -2, -3 and -3') conserved in the MutS family²² are located at the dimer-interface region of domain V (Figs 3c, 5). N-1 and N-3 correspond to the Walker A and B motifs present in many ATPases. N-2 and N-3' are unique to the ABC ATPase superfamily, which includes MutS, UvrA, Rad50, CFTR and ABC transporters²². In the crystal structure of the TAQ MutS–DNA complex, a sulphate ion is hydrogen bonded to the P-loop formed by N-1 (GPNMAGKS, residues 583–590) and marks the ATP-binding site (Fig. 3c). N-3 (DE, residues 662–663) and N-3' (TH, residues 695–696) are located on the β -strands adjacent to the P-loop (Fig. 3c). Interestingly N-2 (ST, residues 636–637), which potentially binds the γ -phosphate, is near the ATP-binding site of the neighbouring subunit, but far from that of the same subunit (Fig. 3c). The crystal structure of a MutS–DNA–ADP ternary complex and mutational studies confirm that N-2 acts in *trans* in ATP hydrolysis (M. S. Junop *et al.*, manuscript in preparation). The composite ATPase active site is supported by the highly conserved MutS dimer interface (Fig. 5) and the observation that

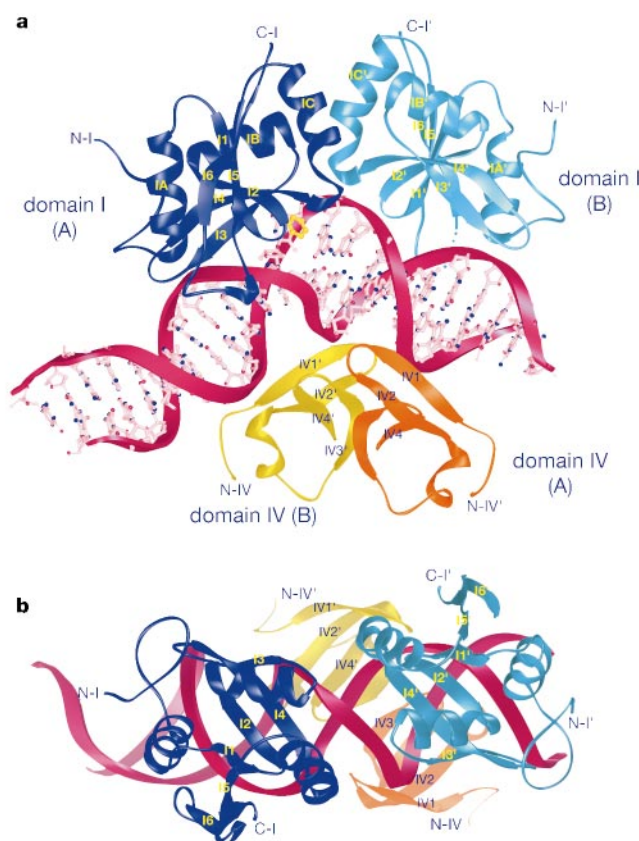
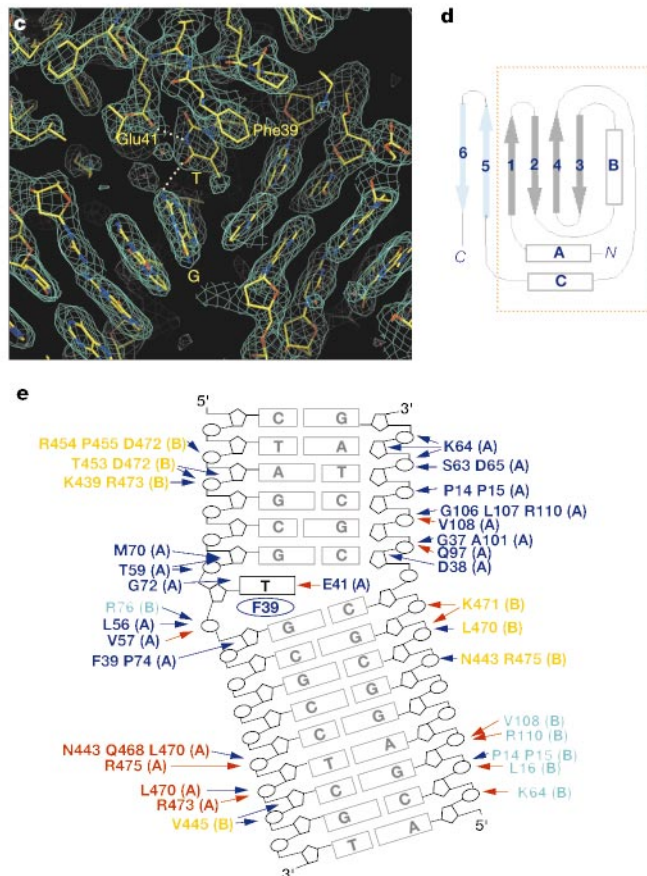


Figure 4 Mismatch recognition by MutS. **a**, The DNA-binding domains represented by ribbon diagrams are shown with the heteroduplex DNA. Domains I and IV of subunit A are shown in dark blue and orange; the corresponding domains of subunit B are in light blue and yellow. The DNA phosphate backbones are represented by red ribbons and sugars and bases are shown as ball-and-stick. Phe 39 of subunit A is shown in yellow. Structural elements of subunit B are denoted with a prime. **b**, A view from the top of **a**. **c**, A $2F_o - F_c$ electron density map contoured at 1.0σ is shown with the refined structure surrounding the unpaired T. The unpaired T and the G 5' to it are labelled. Both Phe 39 and Glu 41



belong to subunit A. The dashed lines represent hydrogen bonds. **d**, Topology diagram of domains I and IV. Features common between the two domains are boxed and shown in grey. αB is replaced by a 3_{10} helix in domain IV. **e**, A diagram of protein–DNA interactions. Only those base pairs that contact MutS are shown. Phosphate groups are represented by ovals between the sugar pentagons. Interactions made by each protein domain are labelled in the same colour as the ribbon diagram of this domain in **a** and **b**. Hydrogen bonds are represented by red arrows and van der Waals contacts by blue arrows.

green, for ATPase activity in red and for interdomain interactions in purple. Residues that, when mutated, cause defective mismatch repair in yeast^{32,36,48–50}, or HNPCC in humans^{3,4} are coloured red, or, if they are highlighted in purple and red, white. The five nucleotide-binding motifs are indicated beneath the sequence alignment.

the ATPase activity depends on a native MutS dimer¹⁵ (P.H. *et al.*, unpublished data).

A composite ATPase active site is probably a common feature among members of the ABC superfamily, which share the conserved nucleotide-binding motifs²² and require dimerization for activity³³. Like MutS, Rad50 also contains a composite ATPase active site³⁴. However, in the crystal structure of HisP, the ATPase subunit of His permease, two protein subunits are packed back-to-back and the conserved nucleotide-binding motif equivalent to N-2 is far away from the bound ATP²³. Perhaps when HisP forms a functional histidine transporter with HisQM subunits, it will dimerize similarly to the dimerization observed in Rad50 and MutS.

Structure prediction of the MutS family

The entire TAQ MutS sequence can be aligned with every member of the MutS family (Fig. 5; and see Fig. 4 in Supplementary Information). The sequence conservation among prokaryotic members indicates that their three-dimensional structures are homologous to that of TAQ MutS (see Fig. 4 in Supplementary Information). The general architecture of TAQ MutS is also conserved among more divergent eukaryotic MutS homologues (Fig. 5). Deletions and insertions in MutS proteins often correlate with the functional diversity. For example, MSH4, MSH5 and Him14 are involved in meiotic DNA segregation but not mismatch repair^{10,35}, and sequences in domain I for mismatch recognition are largely absent in these proteins (Fig. 5). In the absence of domain I, which separates the two channels in TAQ MutS (Fig. 3b), the MSH4–MSH5 dimer is likely to contain a single ~30 Å by 70 Å hole, which should be large enough to encompass two DNA duplexes side by side such as might be encountered during Holliday junction migration.

MutS α and MutS β have different substrate specificity in MMR. MutS α recognizes duplex DNA that contains mispaired or 1–2 unpaired bases, as does TAQ MutS^{1,7}. MutS β has evolved to recognize large unpaired loops, arising with long sequence repeats frequent in eukaryotic genomes. MutS α and MutS β are composed of MSH2–MSH6 and MSH2–MSH3 heterodimers, respectively. On the basis of the asymmetric usage of the two TAQ MutS subunits in mismatch binding, only one subunit in MutS α and MutS β is expected to directly contact mismatched bases. The unique subunit, MSH6 or MSH3, is likely to recognize mismatches as is subunit A of TAQ MutS. All residues important for mismatch recognition by TAQ MutS are conserved in MSH6 (Fig. 5). In MSH3, Lys and Arg residues are found to replace Phe 39 and Glu 41 of TAQ MutS (Fig. 5) and probably interact with the phosphate backbone instead of an unpaired base (Fig. 4c). These substitutions may be important in substrate specificity of MutS α versus MutS β .

The MSH2 subunit lacks the equivalent of Q97 and R110 that are essential for TAQ MutS to anchor domain I in the widened minor groove (Figs 4, 5). It also contains a deletion of 12–14 residues in the region encoding the I β 3–I β 4 hairpin (Fig. 5), which wedges in the widened DNA minor groove and buttresses the nearby Phe 39 in the TAQ MutS–DNA complex (Fig. 4). Mutating the residue equivalent to Phe 39 of TAQ MutS in MSH2 has no effect on mismatch binding by yeast or human MutS α ³¹ (J. Jiricny, personal communication). MSH2 probably plays a role equivalent to that of subunit B of TAQ MutS and stabilizes the kinked conformation of mismatch DNA (Fig. 4). Consistent with this prediction, point mutations in the DNA-binding region of domain IV in yMSH2 cause a mutator phenotype in yeast³⁶ (Fig. 5).

A transmitter connecting ATP- and DNA-binding sites

DNA stimulates the MutS ATPase activity^{14–16}. Conversely, the presence of ATP reduces the affinity of MutS for DNA^{8,14,15}. The crystal structures show that a bound DNA not only immobilizes domains I and IV, but also alters the structures of the ATPase domains (Fig. 2b; and see Fig. 3 in Supplementary Information).

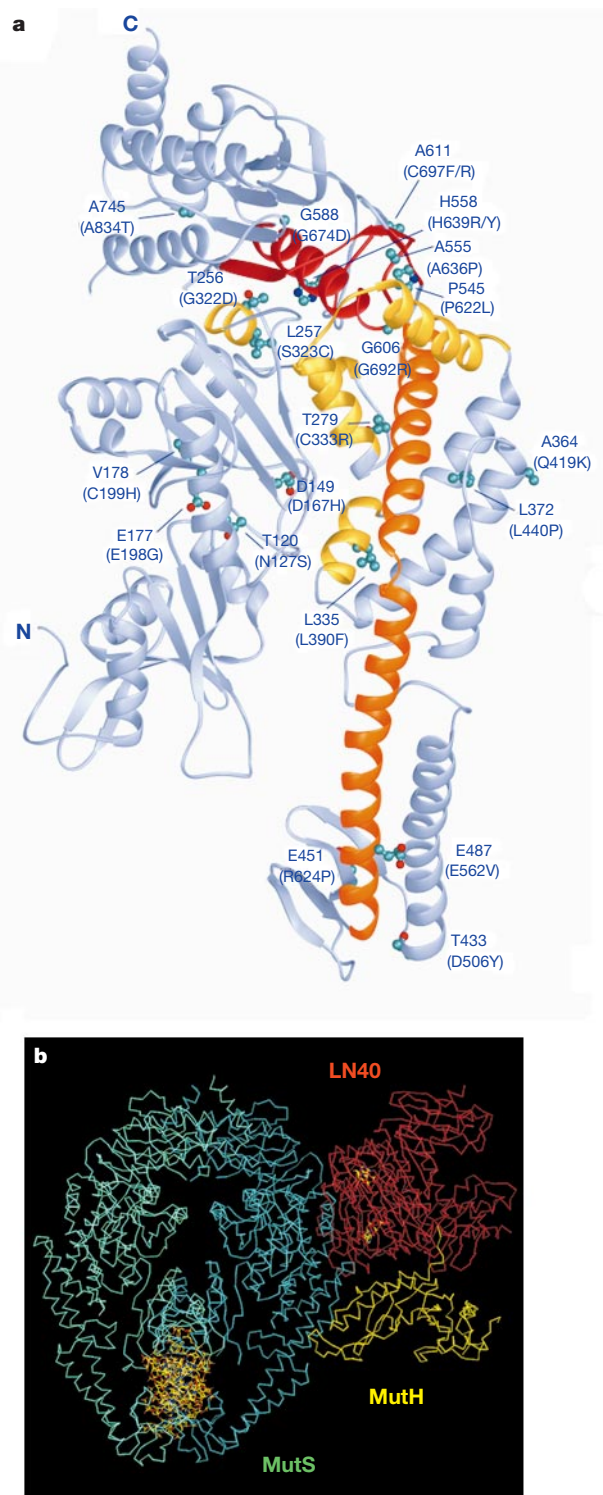


Figure 6 The 'transmitter', HNPCC and MMR. **a**, The ribbon diagram of MutS subunit A. The secondary structures that contain conserved residues to form the 'transmitter' at the three-domain junction are highlighted in red (domain V), yellow (domain III) and orange (III α). IV α B is also highlighted in orange for its role in the allosteric connection. Residues corresponding to the missense mutations found in the HNPCC databases^{3,4} are also shown. Owing to limited clinical and functional data, it is uncertain whether all of these mutations are pathogenic and cause MMR defect. For instance, G322D, L390F and Q419K appear in normal individuals at an allele frequency of 1–2% and might be polymorphisms. However, putative equivalents of these three human MSH2 polymorphisms cause elevated mutation rates in yeast⁴⁸. **b**, A model of the MutS, MutL and MutH complex. The MutS dimer is shown in a blue and green C α trace with the bound heteroduplex DNA, the LN40 dimer in red with two ADPnP molecules, and MutH in yellow. Subunit B of MutS is modelled to contact LN40. The LN40 and MutH interface is based on a previous proposal⁶.

These observations indicate that an allosteric connection exists between the distant ATP- and DNA-binding sites.

An examination of conserved residues in the MutS family reveals the location of a potential allosteric connection. In addition to the ATPase active site and the dimer interface, conserved residues are concentrated at the junction where domains II, III and V converge (Figs 5, 6a). This three-domain junction includes α -helix A and the surrounding β -strands 1, 2, 5 and 6 of domain V, and α -helices A, B, C, D and I of domain III. The functional importance of this region is also indicated by the HNPCC mutations localized here, such as, G322D, S323C, C333R, P622L and H639R/Y (Fig. 6a). Helices V α A, III α I and IV α B form the backbone that directly connects the DNA-binding and the ATPase domains (Figs 2a, 6a). The connection between domains I and III is intervened by the least conserved domain II (Figs 2a, 5) and is probably established only after binding of a heteroduplex DNA. Two HNPCC mutations, N127S and D167H, reside in the I–II domain interface (Fig. 6a), indicating that this interface is critical for MMR. We propose that this three-domain junction serves as a ‘transmitter’ for information exchange between the ATP- and DNA-binding sites.

A model of MutS in mismatch repair

Two previous models propose that MutS leaves the mismatch either by ATP-hydrolysis dependent translocation or ATP-binding dependent sliding before initiating MMR. The bi-directional translocation model based on an α -like DNA-loop structure observed by electron microscopy implied that a heteroduplex DNA passes through a MutS dimer twice⁸. MutS does contain a second channel that by size and electrostatic potential is suitable for DNA binding (Fig. 1). However, a putative MutS dimer of two spheres observed by electron microscopy⁸ actually corresponds to two MutS dimers on the basis of their shape and dimensions. In solution, MutS exists predominantly as a dimer, and formation of tetramers is dependent on protein concentration¹⁹. Both biochemical studies and the crystal structure indicate that it is a MutS dimer that recognizes a mismatch⁹.

The sliding-clamp model was supported by a clamp-like structure of human MutS observed by electron microscopy in the absence of DNA³⁷. Without a bound DNA, domains I and IV are mobile (Fig. 3a). This mobility complicates the interpretation of structural differences shown in the electron microscopy images, which had been attributed to the presence of ATP versus ADP³⁷. The most puzzling problem raised by these two models is how the repair system knows whether and where a mismatch has occurred if MutS leaves a mismatch site.

We propose that MutS remains bound to a mismatch site to faithfully invoke MMR. Our crystal structures suggest that MutS is unable to retain the mismatch-bound conformation once it leaves a mismatch (Fig. 3). Interactions between MutL and MutS have been shown to stabilize the MutS–DNA complex and prevent MutS from leaving a mismatch even in the presence of ATP^{38,39}. Examination of the MutS structure reveals an intriguing depression adjacent to the ‘transmitter’ (see Fig. 5 in Supplementary Information). The two depressions of a MutS dimer differ in shape and electrostatic potential as a result of asymmetric DNA binding (see Fig. 5 in Supplementary Information). We propose that one of the depressions, presumably that of subunit B, the equivalent of MSH2, is the site where MutL interacts with MutS (Fig. 6b) and potentially stabilizes the MutS–DNA interaction. On the basis of the sequence alignment, this depressed surface exists in all members of the MutS family even though the sequence is not conserved. The shape and electrostatic potential of the depression vary, perhaps mirroring a non-conserved surface of MutL proteins. Two HNPCC mutants of human MSH2 (Q419K and L440P) that are exposed to solvent at the outer rim of this depression, mark a potential site for MutS α and MutS β to interact with MutL homologues in humans.

In *E. coli*, MutS that remains at a mismatch site should be able to

activate MutH bound to a GATC sequence either upstream or downstream by protein–protein interactions mediated by MutL, much like enhancer-binding proteins in transcriptional regulation. Our preliminary results also indicate that MutS has to bind both ATP and a mismatch simultaneously to activate MutH and initiate MMR (M. S. Junop *et al.*, manuscript in preparation).

Conclusions

The crystal structures of TAQ MutS and a TAQ MutS–DNA complex capture two states of a flexible and dynamic molecule. They reveal the general architecture of this conserved family of proteins, the basis for recognition of a broad range of mismatches and the existence of a previously unappreciated composite ATPase active site. An allosteric connection between the mismatch- and ATP-binding sites of MutS has emerged from our crystal structures and the biochemical data accumulated over the past decade. The identification of a ‘transmitter’ region together with previous structures of MutL and MutH^{6,40} provide a testable model of the MutS, MutL and MutH complex that initiates MMR in *E. coli*. Finally, the structures and structure-based sequence alignment illuminate the defects of mutations found in various MutS proteins, in particular those of HNPCC kindreds. More experimental data are clearly needed to address whether ATP hydrolysis by MutS is required for activation of MutH through MutL, and to investigate the use of the second channel in the MutS structure that is probably involved in DNA recombination and may also be involved in the search for a mismatch. □

Methods

Crystallization

Full-length TAQ MutS fails to yield diffraction quality crystals of either protein alone or protein–DNA complexes. The truncated TAQ MutS defined by partial proteolysis, amino-acid residues 1–768 and 1–782, were cloned, overexpressed and purified⁹. Selenomethionyl-substituted proteins were expressed in B834 (DE3) cells with defined medium²⁰ and purified over an additional Phenyl-Sepharose column⁹. The oligonucleotides, 5′-GCGACGCTAGCGTGCCTCGTC-3′ and 5′-GGACGAGCCGCGCTAGCGTCG-3′, were synthesized in-house or purchased from Yale University. The MutS–DNA complex was made by incubating MutS (residues 1–768) with the heteroduplex DNA at a protein to DNA molar ratio of 1:1.2 in 20 mM HEPES (pH 7.0), 50 mM KCl, 5 mM MgCl₂, 1.0 mM dithiothreitol and 7% glycerol at 37 °C for 15 min and concentrated in a Centricon-30 from an initial protein concentration of 0.5 mg ml^{−1} to 15 mg ml^{−1}. Crystals of the complex in P2₁2₁2₁ space group (see Table 1 in Supplementary Information) were obtained at 20 °C using the hanging-drop vapour diffusion method. The reservoir solution contained 100 mM cacodylate (pH 6.2), 200 mM (NH₄)₂SO₄, 1 mM dithiothreitol and 18% PEG4000. Crystals appeared within a day and grew to a maximal dimension of ~0.2 × 0.3 × 0.6 mm within a week. Crystals of the TAQ MutS (residues 1–782) in space group P3₁21 were obtained also using the hanging-drop vapour diffusion method. The protein solution was at a concentration of 40 mg ml^{−1} in 20 mM Tris (pH 8.0), 200 mM KCl, 1 mM dithiothreitol, 0.1 mM EDTA, 5 mM MgSO₄ and 5% glycerol, and the reservoir solution contained 90 mM HEPES (pH 7.2), 7.2% PEH400, 1.8 M (NH₄)₂SO₄ and 1.5 mM CuCl₂. Crystals of MutS alone grew at 27 °C and were pyramid-like with the longest dimension up to 0.6 mm.

Data collection

Both types of crystals were flash-frozen in liquid propane for storage and data collection. The cryo-solvent for the MutS–DNA complex contained the mother liquor with 25% PEG4000 and 10% ethylene glycol. The MutS alone crystal was cryo-protected by addition of 15% glucose to the mother liquor. The multiwavelength anomalous diffraction dataset of the MutS–DNA complex at 2.5 Å resolution was collected at four wavelengths with inverse beam geometry at Brookhaven National Laboratory (BNL) beamline X9B using a Quantum4 CCD detector at 100 K (see Table 1 in Supplementary Information). For structure refinement, a dataset of the MutS–DNA complex from a larger crystal of the same selenomethionyl-substituted batch was collected to 2.2 Å resolution in-house at 95 K using a RAXIS-II detector mounted on a Rigaku RU-200 generator. The diffraction data of the MutS alone crystal was collected at BNL beamline X4A at 100 K using an R-Axis IV detector. All diffraction data were processed by HKL (ref. 41) (see Table 1 in Supplementary Information).

Structure determination

SOLVE (ref. 42) found 24 Se sites in the MutS–DNA complex and generated an interpretable electron density map at 2.5 Å (see Table 1 in Supplementary Information). The map was further improved using DM (ref. 43). A complete protein and DNA model were built into the electron density map using O (ref. 44). The resulting model was refined against the 2.2 Å dataset using CNS (ref. 45). The crystal lattice contacts are formed mainly

between MutS molecules with one DNA end packed against a symmetry-related MutS. The final model consists of MutS subunit A (residues 1–628 and 635–765) and B (residues 1–100, 108–628 and 635–762), the DNA heteroduplex, 4 sulphate ions, 7 ethylene glycol and 814 water molecules (see Table 1 in Supplementary Information). In the Ramachandran plot, 91.8% of residues are in the ‘most favourable’ regions and none in ‘disallowed’ regions.

The structure of MutS alone was determined by molecular replacement using CNS. The search model with DNA-binding domains removed gave better solutions than that with the entire MutS. The crystal space group was determined to be $P3_121$ with one MutS dimer in each asymmetric unit. The structure was refined to the R and R_{free} of 33.1% and 36.1%, respectively, with tight NCS restraints (see Table 1 in Supplementary Information). With the B value of the diffraction data near $90 \text{ \AA}^2$ and the low-resolution diffraction, further refinement was judged to be unlikely to provide more information. The final model contains residues 131–197, 203–327, 340–352, 518–628 and 635–765 of one subunit, and residues 117–196, 202–391, 498–628 and 635–762 of the other subunit. In the Ramachandran plot, 98.9% of residues are in the ‘most favoured’ and ‘additionally allowed’ regions and none is in the ‘disallowed’ regions.

Sequence alignment and structure comparison were carried out as described⁴⁰. Figures 4c and 6b were generated using O (ref. 44), Fig. 5 in Supplementary Information was made using GRASP (ref. 46), and the rest of the structure figures were made using RIBBONS (ref. 47).

Received 14 April; accepted 4 August 2000.

- Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65**, 101–133 (1996).
- Umar, A. & Kunkel, T. A. DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur. J. Biochem.* **238**, 297–307 (1996).
- Human Genome Mutation Database [online] (cited 25 March 2000) (<http://www.uwcm.ac.uk/uwcm/mg/ns/1/203983.html>) (1997).
- HNPC Mutation Database [online] (cited 25 March 2000) (<http://www.nfdht.nl/database/msh2.htm>) (1997).
- Herman, J. G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA* **95**, 6870–6875 (1998).
- Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85–97 (1999).
- Jiricny, J. Eukaryotic mismatch repair: an update. *Mutat. Res.* **409**, 107–121 (1998).
- Allen, D. J. *et al.* MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **14**, 4467–4476 (1997).
- Biswas, I. *et al.* Oligomerization of a MutS mismatch repair protein from *Thermus aquaticus*. *JBC* **274**, 23673–23678 (1999).
- Buermeier, A. B., Deschênes, S. M., Baker, S. M. & Liskay, R. M. Mammalian DNA mismatch repair. *Annu. Rev. Genet.* **33**, 533–564 (1999).
- Eisen, J. A. A phylogenomic study of the MutS family of proteins. *Nucleic Acids Res.* **26**, 4291–4300 (1998).
- Matic, I., Taddei, F. & Radman, M. Genetic barriers among bacteria. *Trends Microbiol.* **4**, 69–72 (1996).
- Nakagawa, T., Datta, A. & Kolodner, R. Multiple functions of MutS- and MutL-related hetero-complexes. *Proc. Natl Acad. Sci. USA* **96**, 14186–14188 (1999).
- Gradia, S., Acharya, S. & Fishel, R. The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. *Cell* **91**, 995–1005 (1997).
- Alani, E., Sokolsky, T., Studamire, B., Miret, J. J. & Lahue, R. S. Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Mol. Cell. Biol.* **17**, 2436–2447 (1997).
- Bjornson, K. P., Allen, D. J. & Modrich, P. Modulation of MutS ATP hydrolysis by DNA cofactors. *Biochemistry* **39**, 3176–3183 (2000).
- Su, S. S. & Modrich, P. *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl Acad. Sci. USA* **83**, 5057–5061 (1986).
- Haber, L. T. & Walker, G. C. Altering the conserved nucleotide binding motif in the *Salmonella typhimurium* MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *EMBO J.* **10**, 2707–2715 (1991).
- Biswas, I. & Hsieh, P. Identification and characterization of a thermostable MutS homolog from *Thermus aquaticus*. *J. Biol. Chem.* **271**, 5040–5048 (1996).
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).
- Yang, W. & Steitz, T. A. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**, 131–134 (1995).
- Gorbalenya, A. E. & Koonin, E. V. Superfamily of UvrA-related NTP-binding proteins implication for rational classification of recombination/repair systems. *J. Mol. Biol.* **213**, 583–591 (1990).
- Hung, L. -W. *et al.* Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* **396**, 703–707 (1998).

- Berger, J. M. & Wang, J. C. Recent developments in DNA topoisomerase II structure and mechanism. *Curr. Opin. Struct. Biol.* **6**, 84–90 (1996).
- Jones, M., Wagner, R. & Radman, M. Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* **115**, 605–610 (1987).
- Malkov, V. A., Biswas, I., Camerini-Otero, R. D. & Hsieh, P. Photocross-linking of the NH2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. *J. Biol. Chem.* **272**, 23811–23817 (1997).
- Biswas, I. & Hsieh, P. Interaction of MutS protein with the major and minor grooves of a heteroduplex DNA. *J. Biol. Chem.* **272**, 23811–23817 (1997).
- Bajski, S. R., Jackson, B. A. & Barton, J. K. DNA repair: models for damage and mismatch recognition. *Mutat. Res.* **447**, 49–72 (2000).
- Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399**, 708–712 (1999).
- Dickerson, R. E. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26**, 1906–1926 (1998).
- Bowers, J., Sokolsky, T., Quach, T. & Alani, E. A mutation in the MSH6 subunit of the *Saccharomyces cerevisiae* MSH2-MSH6 complex disrupts mismatch recognition. *J. Biol. Chem.* **274**, 16115–16125 (1999).
- Das Gupta, R. & Kolodner, R. D. Novel dominant mutations in *Saccharomyces cerevisiae* MSH6. *Nature Genet.* **24**, 53–56 (2000).
- Holland, I. B. & Blight, M. A. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J. Mol. Biol.* **293**, 381–399 (1999).
- Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789–800 (2000).
- Zalavsky, J., MacQueen, A. J., Duffy, J. B., Kempthorne, K. J. & Villeneuve, A. M. Crossing over during *Caenorhabditis elegans* meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. *Genetics* **153**, 1271–1283 (1999).
- Studamire, B., Price, G., Sugawara, N., Haber, J. E. & Alani, E. Separation-of-function mutations in *Saccharomyces cerevisiae* MSH2 that confer mismatch repair defects but do not affect nonhomologous-tail removal during recombination. *Mol. Cell. Biol.* **19**, 7558–7567 (1999).
- Gradia, S. *et al.* hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatch DNA. *Mol. Cell* **3**, 255–261 (1999).
- Grilley, M., Welsh, K. M., Su, S. -S. & Modrich, P. Isolation and characterization of the *Escherichia coli* mutL gene product. *J. Biol. Chem.* **264**, 1000–1004 (1989).
- Habraken, Y., Sung, P., Prakash, L. & Prakash, S. ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes. *J. Biol. Chem.* **273**, 9837–9841 (1998).
- Ban, C. & Yang, W. Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* **17**, 1526–1534 (1998).
- Otwonowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. SOLVE: An automated structure solution for MAD and MIR. Edition 1.16 [online] (<http://www.solve.lanl.gov>) (1997).
- CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J. -Y. & Cowan, S. W. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
- Carson, M. Ribbon models of macromolecules. *J. Mol. Graphics* **5**, 103–106 (1987).
- Drotschmann, K., Clark, A. B. & Kunkel, T. A. Mutator phenotypes of common polymorphisms and missense mutations in MSH2. *Curr. Biol.* **9**, 907–910 (1999).
- Drotschmann, K. *et al.* Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. *Proc. Natl Acad. Sci. USA* **96**, 2970–2975 (1999).
- Wu, T. -H. & Marinus, M. G. Dominant negative mutator mutations in the *mutS* gene of *Escherichia coli*. *J. Bacteriol.* **176**, 5393–5400 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank I. Biswas for TAQ MutS expression vectors; Z. Dauter and C. Ogata for synchrotron beamline support; G. Poy for oligonucleotide synthesis; Q. Zhao for assistance in data collection; E. Alani, D. Camerini-Otero, R. Craigie, M. Gellert, M. Junop, T. Kunkel, D. Leahy, K. Mizuuchi, H. Nash and M. Radman for discussions and comments on the manuscript.

Correspondence and requests for materials should be addressed to W.Y. (e-mail: Wei.Yang@nih.gov). Coordinates have been deposited in the Protein Data Bank under accession codes 1EWR (MutS) and 1EWQ (MutS–DNA complex).

The crystal structure of DNA mismatch repair protein MutS binding to a G·T mismatch

Meindert H. Lamers^{*†}, Anastassis Perrakis[‡], Jacqueline H. Enzlin^{*§}, Herrie H. K. Winterwerp^{*}, Niels de Wind^{*||} & Titia K. Sixma^{*}

^{*} Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[‡] European Molecular Biology Laboratory (EMBL), c/o ILL, BP 156, 6 rue Jules Horowitz, 38043 Grenoble, France

[†] These authors contributed equally to this work

DNA mismatch repair ensures genomic integrity on DNA replication. Recognition of a DNA mismatch by a dimeric MutS protein initiates a cascade of reactions and results in repair of the newly synthesized strand; however, details of the molecular mechanism remain controversial. Here we present the crystal structure at 2.2 Å of MutS from *Escherichia coli* bound to a G·T mismatch. The two MutS monomers have different conformations and form a heterodimer at the structural level. Only one monomer recognizes the mismatch specifically and has ADP bound. Mismatch recognition occurs by extensive minor groove interactions causing unusual base pairing and kinking of the DNA. Nonspecific major groove DNA-binding domains from both monomers embrace the DNA in a clamp-like structure. The interleaved nucleotide-binding sites are located far from the DNA. Mutations in human MutS α (MSH2/MSH6) that lead to hereditary predisposition for cancer, such as hereditary non-polyposis colorectal cancer, can be mapped to this crystal structure.

Mismatch repair (MMR) maintains genome stability by correcting mismatches and small insertion or deletion loops (IDLs) introduced by DNA polymerases. In addition, MMR counteracts recombination between homologous but diverged DNA sequences^{1–3}. In *E. coli*, repair is initiated when the dimeric MutS protein recognizes and binds to a mismatch or IDL with an affinity that is only 10–20-fold higher than that of binding to a homoduplex⁴. Almost all mismatches are recognized by MutS and repaired, although there is some variation in the affinity and efficiency of repair depending on the nature of the mismatch and sequence context^{1,4}. Mismatch binding induces ATP uptake and a conformational change in the MutS protein^{5,6}, resulting in a clamp that translocates on DNA^{7,12}. The precise role of ATP hydrolysis in the function of MutS, or its mammalian homologues, is hotly debated. It might act as a motor, providing the energy for bi-directional DNA scanning^{1,7–10}, or might form a switch, signalling between ‘on’ and ‘off’ states to downstream components^{11–13}. Binding of MutS to a mismatch is followed by the ATP-dependent binding^{1,14} of a homodimer of the ATPase MutL^{15–17}. This results in enhancement of bi-directional scanning^{1,7} until a strand-discrimination signal is encountered. In *E. coli*, the signal is a hemimethylated (GATC) sequence that is nicked on the unmethylated (novel) strand by MutH on binding to MutL^{1,15}. Nicking is followed by binding of helicase II, unwinding the DNA towards the mismatch, and the subsequent removal of the mismatch-containing strand using 3'→5' or 5'→3' exonucleases up to discrete sites just beyond the mismatch. Resynthesis and ligation are performed by DNA polymerase III and DNA ligase, in the presence of single-strand binding protein¹.

MMR deficiency and cancer

The initial steps of MMR have been conserved throughout the evolution of eukaryotes; however, the nature of the strand-discrimination signal and downstream steps in mammalian MMR are

largely elusive^{2,3}. In eukaryotes, the heterodimeric MutS homologue MSH2/MSH6 (MutS α) functions in repair of mismatches and short IDLs, whereas MSH2/MSH3 (MutS β) repairs longer IDLs^{2,3}. In addition, the meiosis-specific MSH4/MSH5 heterodimer has diverged to stimulate meiotic crossovers³. The principal MutL homologues, MLH1/PMS2, form a heterodimeric complex (MutL α) which functions in MMR; MLH1 is also involved in stimulating meiotic crossovers. Inherited defects in one allele of MSH2, MLH1, or rarely PMS2 underlie the common cancer predisposition syndrome hereditary non-polyposis colorectal cancer (HNPCC)¹⁷, whereas mutations in MSH6 predispose for cancer with a slightly different phenotype¹⁷. This shows the importance of MMR in preventing accumulation of mutations in cancer-related genes, during the multistage process of tumorigenesis¹⁸.

Structure determination

Here we present the crystal structure of MutS in complex with a 30-base pair (bp) DNA oligomer with a G·T mismatch at position 9. The non-conserved 53 amino acids at the carboxy terminus were truncated (Δ C800). This deletion changes the protein from a tetramer to a dimer *in vitro*. Tetramers were also found for other bacterial MutS proteins at high concentrations^{19,20}, but the expected physiologically relevant form is the dimer. Δ C800 has ATPase activity and DNA mismatch recognition similar to wild type, and can restore MMR activity in a MutS-deficient *E. coli* strain. We solved the structure by a multiwavelength anomalous dispersion (MAD) experiment using seleno-methionine as anomalous scatterer. Refinement against 2.2 Å data resulted in an *R* factor of 22.8% and an *R*_{free} of 26.6%, with excellent stereochemistry (Table 1).

Overall structure

The asymmetric unit contains a MutS dimer bound to one DNA oligomer, with a single G·T mismatch (Fig. 1). The entire MutS protein is well ordered (residues 2–800) except for loop 659–668. Three additional loops are disordered in the non-mismatch-binding monomer (residues 2–13, 57–66 and 95–107). The DNA oligomer is ordered for 16 bp, and 14 bp are invisible.

[§] Present address: Institute for Medical Radiobiology, University of Zürich, Zürich, Switzerland.

^{||} Present address: Radiation Genetics and Chemical Mutagenesis, Leiden University Medical Center, Leiden, The Netherlands.

The MutS dimer is an asymmetric molecule that looks like a pair of praying hands, with the wrists close together, the thumbs coming close and the fingers lightly touching. The DNA is held between thumbs and fingers in this view (Fig. 1). Approximate overall dimensions of the complex are 125×90×55 Å. Almost the entire surface of MutS has an overall negative charge, except for a groove at the ‘top’ of the dimer (Supplementary Information Fig. A). Running through the molecule are several solvent channels, which are lined with negative charges.

MutS is a modular protein (Fig. 2). The amino-terminal mismatch-recognition domain (‘the thumbs’, residues 2–115) forms a six-stranded mixed β-sheet surrounded by three α-helices. A search for similar folds using DALI²¹ reveals that this domain is similar to transfer RNA endonuclease²², superimposing well on the inflexible parts (r.m.s. deviation 2.7 Å, 60 Cα). The connector domain (residues 116–266) is a more regular, mostly parallel β-sheet, surrounded by four α-helices. Its fold resembles the Holliday junction resolvase²³ (ruvC) (r.m.s. deviation 3.0 Å for 113 Cα). The core domain comprises two sequence regions (residues 267–443 and

504–567), which form a helical bundle. From within the core domain, two α-helices (‘the levers’) extend, embracing but not touching the DNA. The core and connector domains together form the ‘palm of the hand’, sharing a large (1,574 Å²) but highly hydrophilic interface. The clamp domain (‘the fingertips’, residues 444–503) is an insertion at the ‘top’ of the levers, forming a four-stranded antiparallel β-sheet. The core domain is connected to the ATPase domain (‘the wrist’, residues 568–765), which has the classical Walker A motif. It is most similar to the ATPase domain of the ABC transporter histidine permease²⁴ (r.m.s. deviation 2.9 Å for 150 Cα). In addition to the normal β-sheet, MutS and His permease have a second β-sheet, which buries the helix succeeding the P-loop (α23). Finally, the helix–turn–helix (HTH) domain (residues 766–800) is involved in dimer contacts (see below).

DNA conformation

The DNA in the structure has mostly B DNA conformation, with a strong kink of ~60° at the G-T mismatch (Fig. 3)²⁵. At the mismatch (base step 8 to 9) the bases do not stack (Fig. 3d) and have distorted

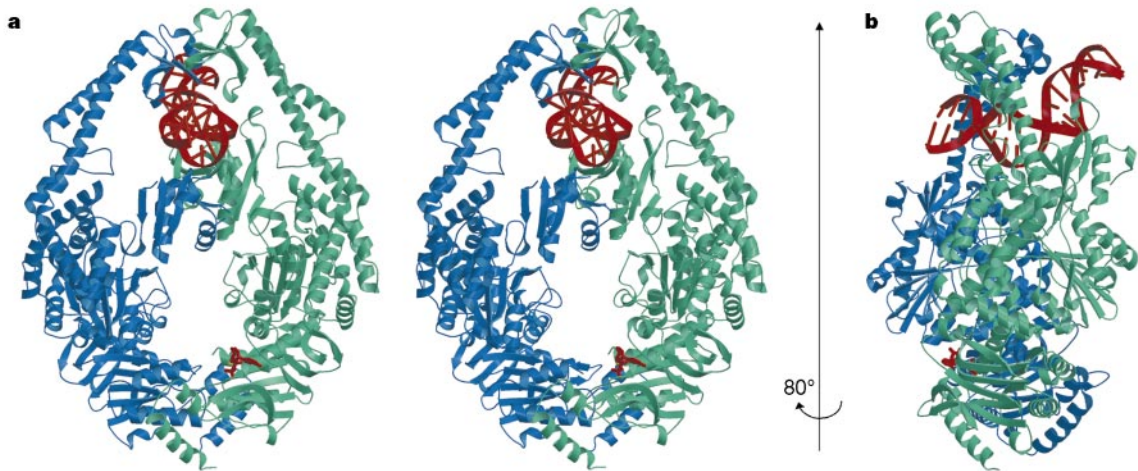


Figure 1 Overview of the MutS–DNA complex. **a**, Stereo view from the front. **b**, Side view, rotated 80° from structure in **a**. DNA and ADP are coloured red, the mismatch-binding monomer light green, and the second monomer blue.

Table 1 Data collection and refinement statistics

Data set		λ_1 (peak)		λ_2 (inflection)		λ_3 (remote)		High resolution
λ (Å)		0.9788		0.9794		0.8855		0.933
Resolution (Å)		3.0 (3.1–3.0)		3.0 (3.1–3.0)		2.7 (2.8–2.7)		20–2.2 (2.24–2.20)
R_{merge} (%)		5.3 (17.6)		5.4 (17.2)		7.8 (47.4)		6.5 (46.9)
$I/\sigma I$		22.5 (4.9)		21.9 (4.9)		15.2 (1.9)		15.9 (2.2)
Redundancy		3.6 (1.9)		3.6 (1.9)		3.5 (1.9)		3.7 (3.1)
Completeness		99.4 (94.2)		99.4 (94.2)		99.9 (89.7)		98.7 (95.1)
Phasing		<u>Iso</u>	<u>Ano</u>	<u>Iso</u>	<u>Ano</u>	<u>Iso</u>	<u>Ano</u>	
R_{cullis} %		0.36	0.49	0.38	0.64	n.a.	0.74	
Phasing power		5.1	4.3	4.9	3.4	n.a.	2.8	
Refinement								
Reflections	109,667							
R_{free} refl.	2,199							
R_{p}	22.8 (32.4)							
R_{free} %	26.6 (36.7)							
No. of atoms								
Total	13,392							
Protein	12,171							
DNA	714							
Ligands	37							
Water	470							
r.m.s. bonds (Å)	0.012							
r.m.s. angles (°)	1.173							

All numbers in parentheses refer to the high-resolution shell. Statistics for the three MAD data sets are for unmerged Bijvoet pairs. All data in high-resolution data set used for refinement. n.a., not applicable.

orientations, as shown by large roll, tilt and twist angles. The DNA has adapted to this kink by adjustments in the sugar pucker, from C2'-endo (observed mainly in B DNA) to C3'-endo (characteristic for A DNA) for six nucleotides (Fig. 3e). This results in a very wide and shallow minor groove and an associated narrow and deep major groove for base pairs 8–11. The DNA is not involved in crystal packing.

Previous crystal structures²⁶ of free DNA with a G-T mismatch do not show similar macroscopic changes. In these structures, as well as in the structure of the complex with G-T specific glycosylase Vsr²⁷, the G-T mismatched bases form a wobble base pair. A completely different hydrogen-bonding pattern is observed in the MutS-bound mismatched base pair (Fig. 3f).

Nonspecific DNA interactions

The clamp domains in both monomers contact the DNA backbone in a sequence-independent manner through limited protein–DNA interfaces (230 Å² and 291 Å²), with interactions extending over 13 bp (Fig. 3b, c, e). The non-mismatch-binding monomer spans the major groove surrounding the mismatch. This is facilitated by the kink in the DNA and the narrow major groove. The clamp domains present a positively charged surface to the DNA, forming several salt bridges with the phosphate backbone (Fig. 3e; Supple-

mentary Information Fig. B) and might thus function in initial recognition of homoduplex DNA by MutS.

Mismatch recognition

Only one monomer has specific mismatch-binding contacts, involving exclusively its N-terminal mismatch-recognition domain. The equivalent domain in the second monomer has only loose DNA backbone contacts elsewhere (Fig. 3e; Supplementary Information Fig. B). The interactions between DNA and the mismatch-binding domain are distributed over a large surface area (1,250 Å²) on the minor groove side of the DNA (Fig. 3g). Phe 36 wedges into the DNA, stacking with the mismatched thymine (thymine 22) (Fig. 3d). The importance of the equivalent phenylalanine in *Thermus aquaticus* MutS was noted previously, and mutation to Ala markedly reduced DNA binding²⁸. The bases of the mismatch are recognized by van der Waals contacts from strand β 4 (residues 68–70), two main-chain hydrogen bonds and the Oe2 of the side chain of Glu 38 (Fig. 3f). The latter residue forms a bifurcated hydrogen bond to guanine 10 as well as thymine 22 of the mismatch. Two negatively charged protein side chains (Asp 35 and Glu 38) are centrally located in the middle of the widened minor groove (Fig. 3g). In a normal minor groove these residues would cause electrostatic repulsion of the DNA backbone, but in the widened minor groove they may help to orientate the mismatch-binding domain with respect to the DNA backbone. The only additional contact to a base (Arg 58 to O2 of cytosine 11) requires no base specificity. Thus, the MutS protein is probing for the mismatch and DNA deformability primarily, and additional sequence context specificity must be conferred by structural flexibility rather than specific interactions with bases.

The only difference between the N-terminal domains in the two monomers is the flexibility of three loops in the second monomer (r.m.s. deviation, 0.56 Å for 79 C α). Thus, most of the mismatch-recognition surface is preformed, scanning the DNA with Phe 36, Asp 35 and Glu 38 already orientated. On mismatch recognition, an induced fit of the three flexible loops should occur to form the full binding surface.

There is no overall positive charge on the mismatch-binding domain (Fig. 3g), leaving initial DNA recognition to the clamp domains. Once bound, MutS will continuously try to bend the DNA through the combined action of the mismatch-recognition domain and the clamp domains. The preformed part of the mismatch-binding site will attempt to modulate the DNA conformation through its negative charges and to wedge in the phenylalanine. Formation of the complete binding site will only succeed where there is a flexible DNA backbone owing to a mismatch or looped out base, or where a bend is already present in the DNA owing to chemical modification by an intrastrand cisplatin adduct or other DNA adducts recognized by MutS²⁹.

In the human heterodimers MSH2/MSH3 and MSH2/MSH6, mismatch discrimination is dependent on MSH3 or MSH6 but not on MSH2 (refs 2, 3). Thus, the mismatch-binding monomer in the MutS structure most probably behaves as MSH3 or MSH6, whereas the other monomer assumes the role of MSH2. Relative to the MSH6-containing dimer the MSH2/MSH3 heterodimer is less efficient in recognizing base–base mismatches^{2,3}. MSH3 has lost two residues contacting the mispaired G-T (Phe 36 and Glu 38; Supplementary Information Fig. B). Both are converted to basic residues, markedly changing the binding surface. Human MSH2 has lost many of the DNA-contacting residues, including the side chains that hydrogen bond to the DNA (Fig 3e; Supplementary Information Fig. B).

Dimer interaction and domain movement

The two MutS monomers interact with each other through two regions—the clamp domains (residues 444–503) and the C-terminal ATPase and HTH domains (residues 568–800). The two clamp

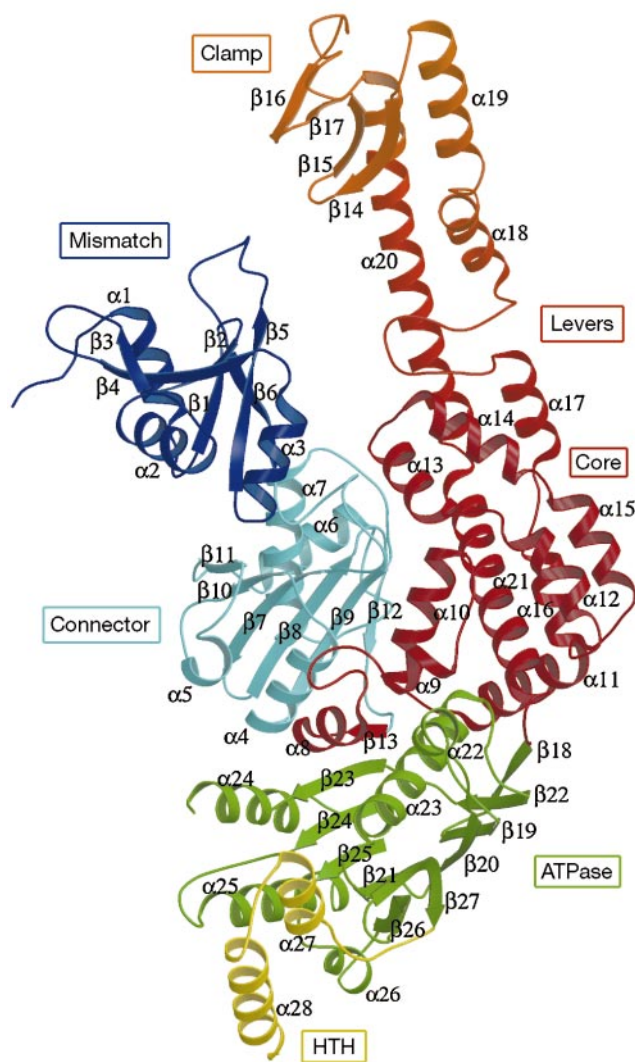


Figure 2 Mismatch-binding monomer coloured by domain. The mismatch-binding domain (2–115) is coloured dark blue, the connector domain (116–266) light blue, the core domain and levers (267–443 + 504–567) from red to orange, the clamp domain (444–503) orange, the ATPase domain (568–765) green, and HTH (766–800) yellow.

domains have limited interactions (Fig. 3a–c; Supplementary Information Fig. B), sharing only 451 Å² contact surface and two intramolecular hydrogen bonds. An excess of positive charges exists in the interface, which could prevent tight clamp formation in the

absence of DNA. For docking onto DNA, the clamp domain dimer has to open. Once DNA is bound, the charges are partially compensated and the observed interaction can take place. Further rearrangements resulting in a tighter clamp are possible on, for

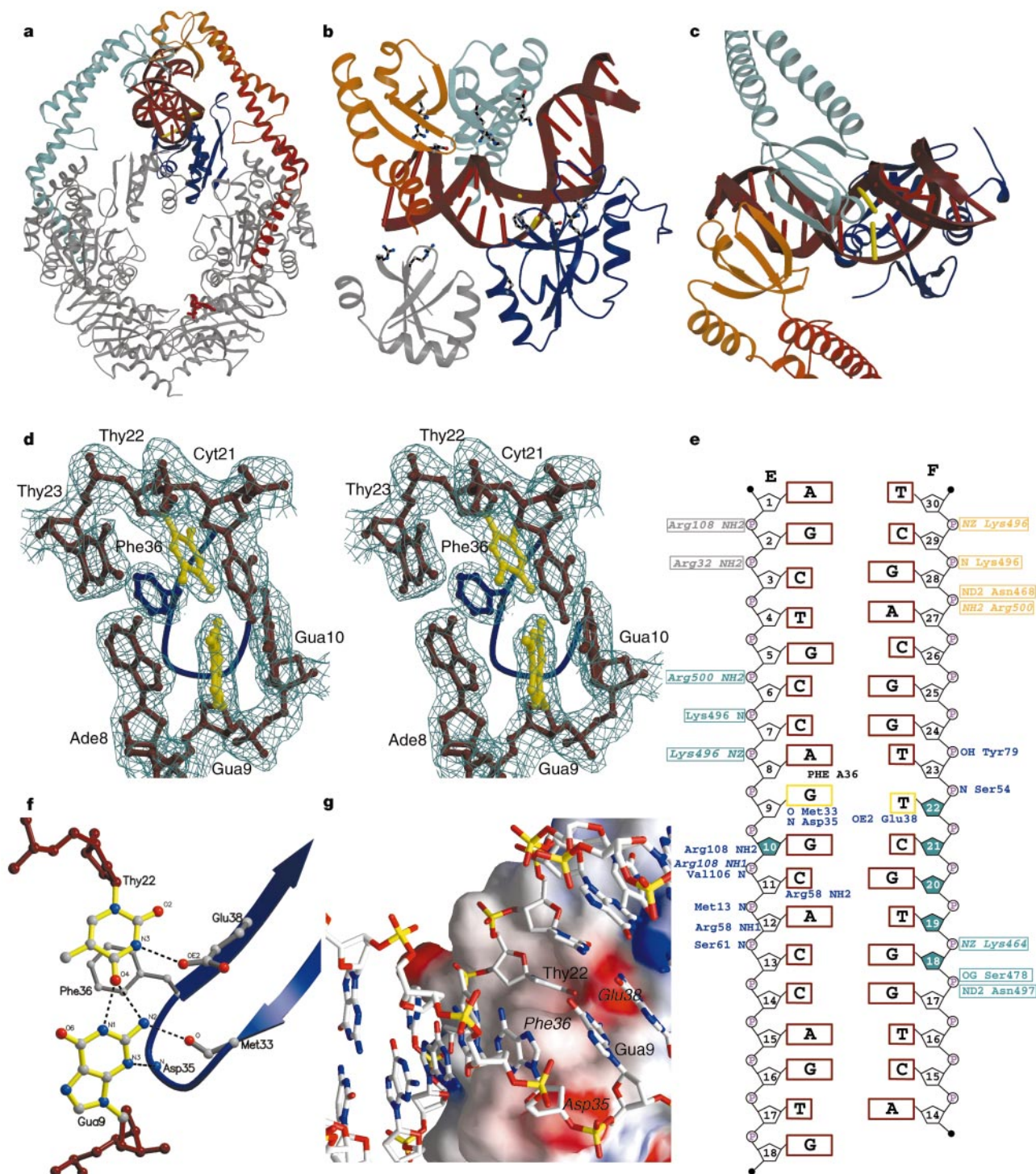


Figure 3 Interaction of MutS with mismatched DNA. **a**, Overview showing MutS embracing the DNA mismatch-binding domain in dark blue, its lever and clamp domains in red to orange, and lever and clamp domains of the second monomer in light blue. DNA and ADP are coloured red, the rest of the structure is grey. **b**, Side view, showing all residues involved in hydrogen bonds and salt bridges to DNA. Mismatch is coloured yellow. **c**, Top view, showing the asymmetric binding of the clamp domains. **d**, Stereo view of the sigma A-weighted ($2m_F - DF_c$)⁴³ electron density map contoured at 1.0 r.m.s. displaying stacking of Phe 36 (dark blue) on the mismatched thymine 22 (yellow). **e**, Diagram of all hydrogen bonds and salt bridges (in italics) made to DNA.

Residues are coloured by domain as in **a**; nonspecific contacts are boxed. Sugar puckering of C3' endo (A form DNA) is indicated by light blue colouring of riboses. **f**, Mismatch-specific hydrogen bonding. The G-T mismatch does not pair as a wobble base, which would have hydrogen bonds between guanine O6 and thymine N3, and guanine N1 and thymine O2. **g**, Electrostatic surface representation⁴⁸ of mismatch-recognition domain in contact with DNA (stick model), displaying the lack of positive charge in the interface. Electrostatic charges (red, negative; blue, positive) are shown between -10 and +10 kT. Mismatched bases are labelled, as well as the approximate surface area of Phe 36, Asp 35 and Glu 38.

example, ATP binding.

The MutS dimer interface in the ATPase domain is extensive (Fig. 4b; Supplementary Information Fig. B), burying 2,922 Å². Two bundles of four helices are formed on opposite sides of the dimer, by the HTH region of each monomer packing against helices α24 and α25 of the other monomer (Fig. 4b). Further interleaving of the domains occurs through reciprocal contacts of a loop (between β25 and α26) with the N terminus of helix α25 of each monomer. The C termini of the truncated protein are close to the dimer interface, and it is possible that the interface in the full-length protein is even larger.

Comparison of the conformation of the two MutS monomers by superimposing their ATPase domains shows the asymmetry of the dimer (Supplementary Information Fig. C). Although all domains are reorientated slightly, analysis of the movements with Dyndom³⁰ shows that a 'central' region can be defined, consisting of the core, connector and ATPase domains, which is rotated by 177° between the two monomers. Relative to this rearrangement, the mismatch recognition, clamp and HTH domains are reorientated to varying degrees (by 29°, 14° and 9°, respectively).

Nucleotide binding and hydrolysis cycle

The nucleotide-binding sites are buried in the dimer interface. Density for ADP and Mg²⁺ is present only in the mismatch-binding

monomer. The P-loop conformation of the second monomer needs to rearrange before it can bind nucleotide. MutS binds ADP in the classical P-loop mode (Fig. 4a). The adenine group stacks between His 760 and Phe 596. Adenine specificity is conferred by two hydrogen bonds with main-chain atoms of Ile 597. A Mg²⁺ ion is octahedrally coordinated to the β-phosphate, the hydroxyl group of Ser 621 and four water molecules. The Walker B motif (D-E-x-x, residues 693–696) is involved in stabilizing Mg²⁺-bound water.

The structure reveals possible catalytic residues. The side chain of Asn 616 is in a similar position to the catalytic glutamate in G proteins³¹ and might have a similar role in MutS. The conserved His 728 could be involved in destabilizing the γ-phosphate, but this would require reorientation of its χ1 dihedral angle to the position seen in the ABC transporter²⁴. This reorientation is prevented by a dimer contact (loop 698–701) in the current MutS structure. The disordered loop (659–668) is highly conserved and might be involved in catalysis. Conserved residues Asp 661 and Ser 668 could reach the active site of either monomer. Rearrangements on ATP binding could bring these or additional residues closer to the phosphate, enabling catalysis.

Extrapolating from other P-loop NTPases the conformation of several regions in MutS could change on ATP binding and hydrolysis (Fig. 4b, c). The 'lid' region of adenylate kinase³² is involved in ATP-dependent movement. The equivalent area in MutS (residues

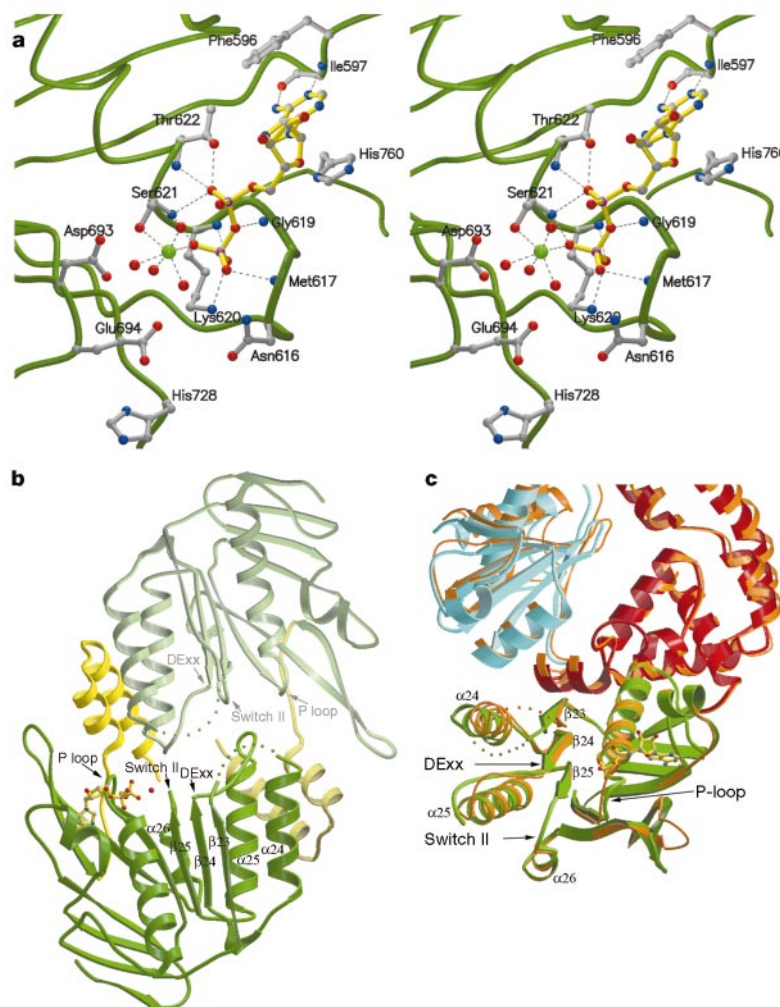


Figure 4 ATPase domain of MutS. **a**, Stereo view of ADP binding showing the ADP-contacting residues in grey, the ADP molecule in yellow, Mg²⁺ in bright green, and MutS in dark green. **b**, ATPase domains and dimer interface viewed from the top. Mismatch-binding monomer below (ATPase domain, green; HTH yellow), second monomer above (ATPase domain, pale green; HTH, pale yellow). Residues 659–668 are disordered in

both molecules (dotted lines). **c**, Superposition of the ATPase domains showing changes in conformation. The ATPase domain of the mismatch-binding monomer is coloured green, the core domain red, the connector domain light blue, and the superimposed second monomer orange.

742–800) is involved in dimer contacts. In GTPases³¹, switches I and II show the main ATP-dependent movements. The highly conserved flexible loop (residues 659–668) is not a topological equivalent for switch I, but it could occupy a similar site, coming from either monomer. This loop could contact the connector domain (within the same monomer) if it acquires the conformation observed in the equivalent (but longer) loop in the ABC transporter histidine permease²⁴. In any case the loop will contact the ATPase domain of the opposite monomer, possibly directly to the opposing active site. The switch II equivalent region in MutS (the loop between strand β 25 and helix α 26) is also involved in dimer contacts (Fig. 4b). Thus, all three regions expected to show ATP-dependent movements would act through the dimer interface, possibly involving allosteric effects.

DNA scanning and signal transduction

Our finding of a single ADP molecule, bound in the mismatch-binding monomer, fits with data showing that the two ATPase sites are non-equivalent^{10,33,34}. It also shows that even the homodimeric MutS protein behaves as a structural heterodimer, similar to its eukaryotic homologues. As biochemical data indicate that communication occurs between the DNA-recognition and ATPase domains^{1–3,8–13,33–35}, the observed ADP distribution could be the result of mismatch binding. It may indicate increased affinity for ADP in the mismatch-binding monomer, or an induced release of ADP in the second monomer.

Subsequently, it is likely that the other (non-mismatch-binding) monomer will proceed with ATP uptake. On ATP binding, MutS is known to release from a mismatch, while remaining bound to the DNA^{1–3,8–13,35}. After analysis of the structure this release is surprising, because of the large surface area involved in mismatch recognition and the extensive deformation in the DNA. However, most experiments showing DNA release have been performed in the absence of MutL. ATP-dependent release in the presence of MutL is less obvious^{14,36}. Thus, it is possible that MutL stabilizes the binding of MutS to the mismatch. Such a mechanism could increase the efficiency of mismatch discrimination from the approximate 10-fold discrimination of mismatched DNA over homoduplex observed for MutS alone⁴. Subsequently, the ternary MutS–MutL complex^{1,15,16} will search for a strand-discrimination signal.

We would like to propose that mismatch binding is retained during the search for a strand-discrimination signal. One possibility is the involvement of the observed MutS tetramer^{10,19,20}, with one dimer binding the mismatch and the other searching for the strand-discrimination signal. Another possibility is direct binding of the MutS–MutL mismatch complex to the strand-discrimination complex. Alternatively, MutS could use the nonspecific DNA binding of the clamp to thread DNA through the eye of the (clamp) needle, while the mismatch-binding domain remains tethered to the mismatch, thus allowing α -loop formation. On threading the DNA through the clamp, rearrangements of domains within the centre of the protein will be required. The hydrophilic nature of the interface between the core and connector domains would allow this opening up to reveal a more positively charged surface.

HNPCC mutations

Some of the MSH2 or MSH6 mutations found in families with cancer predisposition such as HNPCC are missense variants, involving the change of a single amino acid. These mutations have tentatively been categorized as pathogenic or polymorphisms in databases on the World-Wide Web (Supplementary Information Fig. D). However, the limited amount of functional and genetic data prevents definitive classification of individual variants as pathogenic (causing a MMR defect) or polymorphism. Because additional functional studies of these variants are needed, it is of interest to analyse the location of these mutations in our structure. This requires mapping the human variants to the equivalent

residues in *E. coli* MutS (Supplementary Information Fig. B)⁵⁰. For detailed analysis critical discussion of the local quality of the alignment will be necessary (M.H.L., N.W. and T.K.S., manuscript in preparation). However, current mapping of the missense mutations on the MutS structure (Supplementary Information Fig. D) does show that they are widely distributed over the structure, indicating the importance of, for example, the connector and clamp domains. Mutations occur both in solvent-exposed and buried positions, including several in interdomain contacts. Some mutations are found in interfaces involved in nonspecific DNA binding, dimerization and ATP binding. Thus, the pathogenic mutations reflect the modular multifunctional nature of this complex protein. □

Methods

Purification

E. coli MutS (853 amino acids) and Δ C800 (1–800) were expressed from plasmids pMQ372 (ref. 37) and pM800, respectively, in B834(DE3)pLysS using fortified minimal medium for seleno-methionine substitution. The proteins were purified using a modified procedure³⁸ with ion exchange (MonoQ and heparin) and gel-filtration chromatography (Superdex 200) as final steps. MutS was concentrated to ~ 14 mg ml^{−1} in 25 mM HEPES, pH 7.5, 250 mM NaCl, 10–20 mM β -mercaptoethanol.

Protein analysis

ATPase assays were carried out spectrophotometrically³⁹, and DNA band-shift experiments were carried out using established protocols¹⁰. Reconstitution of MMR activity was scored by monitoring rifampin resistance³⁷ in MutS-deficient *E. coli* strain RK1517 (MutS201::Tn5)⁴⁰.

Crystallization

MutS was crystallized in a 2:1 molar ratio with a synthetic 30-nucleotide oligonucleotide containing a G-T mismatch at position 9 (top strand, 5'-AGC TGC CAG GCA CCA GTG TCA GCG TCC TAT -3'; bottom strand, 3'-TCG ACG GTT CGT GGT CAC AGT CGC AGG ATA-5'; mismatch underlined, residues not defined in the structure in italics). We grew crystals at room temperature using microseeding in hanging drops with a well solution containing 12–14% PEG 6000, 150–300 mM NaCl, 100 mM HEPES pH 7–8, 10 mM MgCl₂ and 100–150 μ M ADP. Crystallization only occurred after C-terminal degradation or with the Δ C800 protein. The space group is P2₁2₁2₁; the cell dimensions are $a = 89.96$ Å, $b = 92.37$ Å, $c = 261.33$ Å.

Data collection and analysis

MAD diffraction data were collected to 3.0 Å resolution at the ESRF synchrotron in Grenoble from a single seleno-methionine crystal at 100 K on beamline BM14. High-resolution data were collected to 2.2 Å on beamline ID14/EH2 from the same crystal. All data were processed with the HKL2000 package⁴¹. Selenium sites were located by direct methods using SNaB⁴² or SHELXM (G. Sheldrick, personal communication). Initial phasing was done with Mlphare⁴³ and final phase probability calculations with SHARP⁴⁴. Density modification by DM⁴⁵ yielded an excellent map, and a model of one monomer was built in O⁴⁶. We used molecular replacement to position the second monomer, which was refined as several rigid bodies and manually adjusted in O. Non-crystallographic symmetry restraints were not used for refinement, owing to the substantial differences observed between the two monomers. Refinement was performed with CNS⁴⁶ and REFMAC⁴⁷. Details of the crystallographic analysis are in Table 1. We carried out electrostatic surface calculations with Grasp⁴⁸, and sequence alignment figure (Supplementary Information Fig. B) and conservation analysis (Supplementary Information Fig. D) with Espript⁴⁹. We used the following cut-offs for inclusion of contacts in figures: van der Waals interactions, 3.8 Å; salt bridges, 4.0 Å; hydrogen bonds, 3.5 Å. Maximum donor/H/acceptor angles were 60°; hydrogen/acceptor/next atom angles were 90°.

Received 5 May; accepted 18 August 2000.

- Modrich, P. & Lahue, R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65**, 101–133 (1996).
- Kolodner, R. D. & Marsischky, G. T. Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* **9**, 89–96 (1999).
- Buermeyer, A. B., Deschenes, S. M., Baker, S. M. & Liskay, R. M. Mammalian DNA mismatch repair. *Annu. Rev. Genet.* **33**, 533–564 (1999).
- Su, S. S., Lahue, R. S., Au, K. G. & Modrich, P. Mismatch specificity of methyl-directed DNA mismatch correction *in vitro*. *J. Biol. Chem.* **263**, 6829–6835 (1988).
- Joshi, A., Sen, S. & Rao, B. J. ATP-hydrolysis-dependent conformational switch modulates the stability of MutS-mismatch complexes. *Nucleic Acids Res.* **28**, 853–861 (2000).
- Biswas, I. & Vijayargia, R. Heteroduplex DNA and ATP induced conformational changes of a MutS mismatch repair protein from *Thermus aquaticus*. *Biochem. J.* **347**, 881–886 (2000).
- Allen, D. J. *et al.* MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **16**, 4467–4476 (1997).
- Blackwell, L. J., Bjornson, K. P. & Modrich, P. DNA-dependent activation of the hMutS α ATPase. *J. Biol. Chem.* **273**, 32049–32054 (1998).
- Blackwell, L. J., Martik, D., Bjornson, K. P., Bjornson, E. S. & Modrich, P. Nucleotide-promoted release

- of hMutSα from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. *J. Biol. Chem.* **273**, 32055–32062 (1998).
10. Bjornson, K. P., Allen, D. J. & Modrich, P. Modulation of MutS ATP hydrolysis by DNA cofactors. *Biochemistry* **39**, 3176–3183 (2000).
 11. Gradia, S., Acharya, S. & Fishel, R. The human mismatch-recognition complex hMSH2–hMSH6 functions as a novel molecular switch. *Cell* **91**, 995–1005 (1997).
 12. Gradia, S. *et al.* hMSH2–hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol. Cell* **3**, 255–261 (1999).
 13. Gradia, S., Acharya, S. & Fishel, R. The role of mismatched nucleotides in activating the hMSH2–hMSH6 molecular switch. *J. Biol. Chem.* **275**, 3922–3930, (2000).
 14. Galio, L., Bouquet, C. & Brooks, P. ATP hydrolysis-dependent formation of a dynamic ternary nucleoprotein complex with MutS and MutL. *Nucleic Acids Res.* **27**, 2325–2331 (1999).
 15. Ban, C. & Yang, W. Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* **95**, 541–552 (1998).
 16. Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85–97 (1999).
 17. Lynch, H. T. & De La Chapelle, A. Genetic susceptibility to non-polyposis colorectal cancer. *J. Med. Genet.* **36**, 801–811 (1999).
 18. Vogelstein, B. & Kinzler, K.W. The multistep nature of cancer. *Trends Genet.* **9**, 138–141 (1993).
 19. Takamatsu, S., Kato, R. & Kuramitsu, S. Mismatch DNA recognition protein from an extremely thermophilic bacterium, *Thermus thermophilus* HB8. *Nucleic Acids Res.* **24**, 640–647 (1996).
 20. Biswas, I. *et al.* Oligomerization of a MutS mismatch repair protein from *Thermus aquaticus*. *J. Biol. Chem.* **274**, 23673–23678 (1999).
 21. Holm, L. & Sander, C. Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* **25**, 231–234 (1997).
 22. Li, H., Trotta, C. R. & Abelson, J. Crystal structure and evolution of a transfer RNA splicing enzyme. *Science* **280**, 279–284 (1998).
 23. Ariyoshi, M. *et al.* Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell* **78**, 1063–1072 (1994).
 24. Hung, L. W. *et al.* Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* **396**, 703–707 (1998).
 25. Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655–667 (1989).
 26. Hunter, W. N. Crystallographic studies of DNA containing mismatches, modified and unpaired bases. *Methods Enzymol.* **211**, 221–231 (1992).
 27. Tsutakawa, S. E., Jingami, H. & Morikawa, K. Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* **99**, 615–623 (1999).
 28. Malkov, V. A., Biswas, I., Camerini-Otero, R. D. & Hsieh, P. Photocross-linking of the NH₂-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. *J. Biol. Chem.* **272**, 23811–23817 (1997).
 29. Marra, G. & Schar, P. Recognition of DNA alterations by the mismatch repair system. *Biochem. J.* **338**, 1–13 (1999).
 30. Hayward, S. & Berendsen, H. J. C. Systematic analysis of domain motions in proteins from conformational change; new results on citrate synthase and T4 lysozyme. *Proteins Struct. Funct. Genet.* **30**, 144–154 (1998).
 31. Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).
 32. Gerstein, M., Schulz, G. & Chothia, C. Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers. *J. Mol. Biol.* **229**, 494–501 (1993).
 33. Iaccarino, I., Marra, G., Palombo, F. & Jiricny, J. hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSα. *EMBO J.* **17**, 2677–2686 (1998).
 34. Studamire, B., Quach, T. & Alani, E. *Saccharomyces cerevisiae* Msh2p and Msh6p ATPase activities are both required during mismatch repair. *Mol. Cell. Biol.* **18**, 7590–7601 (1998).
 35. Iaccarino, I., Marra, G., Dufner, P. & Jiricny, J. Mutation in the magnesium binding site of hMSH6 disables the hMutSα sliding clamp from translocating along DNA. *J. Biol. Chem.* **275**, 2080–2086 (2000).
 36. Habraken, Y., Sung, P., Prakash, L. & Prakash, S. ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2–MSH6 and MLH1–PMS1 protein complexes. *J. Biol. Chem.* **273**, 9837–9841 (1998).
 37. Wu, T. H. & Marinus, M. G. Deletion mutation analysis of the mutS gene in *Escherichia coli*. *J. Biol. Chem.* **274**, 5948–5952 (1999).
 38. Su, S. S. & Modrich, P. *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad. Sci. USA* **83**, 5057–5061 (1986).
 39. Panuska, J. R. & Goldthwait, D. A. A DNA-dependent ATPase from T4-infected *Escherichia coli*. Purification and properties of a 63,000-dalton enzyme and its conversion to a 22,000-dalton form. *J. Biol. Chem.* **255** 5208–5214 (1980).
 40. Fishel, R. A., Siegel, E. C. & Kolodner, R. Gene conversion in *Escherichia coli*. Resolution of heteroallelic mismatched nucleotides by co-repair. *J. Mol. Biol.* **188**, 147–157 (1986).
 41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
 42. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Crystallogr.* **32**, 120–124 (1999).
 43. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 44. La Fortelle, E. de & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472–494 (1997).
 45. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 46. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
 47. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
 48. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
 49. Gouet, P., Courcelle, E., Stuart, D. I. & Metz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
 50. Culligan, K. M., Meyer-Gauen, G., Lyons-Weiler, J. & Hays, J. B. Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. *Nucleic Acids Res.* **28**, 463–471 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Marinus for MutS plasmid pMQ372 and purification protocols; P. Modrich for strain RK1517; S. Cusack and H. te Riele for support and enthusiasm; J. Tainer for useful suggestions; G. Sheldrick for discussions on twinning; C. Vornrhein for assistance with SHARP; K. Culligan and J. Hays for use of their alignment; and staff at the ESRF and EMBL outstations Hamburg and Grenoble, in particular G. Leonard, for support in data collection. This work was supported by the Dutch Cancer Society and an EMBO post-doctoral fellowship to A.P., and by European Union TMR/LSF grants for data collection visits.

Correspondence and requests for materials should be addressed to T.K.S. (e-mail: sixma@nki.nl). Coordinates have been deposited in the protein data bank under accession number 1E3M.

Dynamics of singularities in a constrained elastic plate

Arezki Boudaoud*, Pedro Patrício†, Yves Couder* & Martine Ben Amar*

* Laboratoire de Physique Statistique de l'ENS (associated with CNRS and the universities Paris VI and Paris VII), 24 rue Lhomond, F-75231 Paris Cedex 05, France

† Centro de Física da Matéria Condensada, Universidade de Lisboa Av. Prof. Gama Pinto 2, 1649-003 Lisbon, Portugal

Large deformations of thin elastic plates usually lead to the formation of singular structures which are either linear^{1–4} (ridges) or pointlike^{5–8} (developable cones). These structures are thought to be generic for crumpled plates^{3,5}, although they have been investigated quantitatively only in simplified geometries^{1–4,6–8}. Previous studies^{9–11} have also shown that a large number of singularities are generated by successive instabilities. Here we study, experimentally and numerically, a generic situation in which a plate is initially bent in one direction into a cylindrical arch, then deformed in the other direction by a load applied at its centre. This induces the generation of pairs of singularities; we study their position, their dynamics and the corresponding resistance of the plate to deformation. We solve numerically the equations describing large deformations of plates; developable cones are predicted, in quantitative agreement with the experiments. We use geometrical arguments to predict the observed patterns, assuming that the energy of the plate is given by the energy of the singularities.

In the absence of constraints, the two main curvatures κ_1 and κ_2 of a plane plate are zero. It is easy to bend the plate, that is, to give it a finite curvature in one direction so that it becomes cylindrical or conical. The curvature and thus the bending energy are then approximately evenly spread in the plate. It scales as $E_b \approx Eh^3$, E being the Young's modulus of the plate material and h its thickness. However, once bent in one direction, it is difficult to bend the plate also in the other direction because this will tend to create a finite Gauss curvature ($G = \kappa_1\kappa_2$) which can be only be acquired by stretching. The stretching energy is $E_s \approx EhR^2$, R being a typical length of the plate. As $E_s/E_b \approx (R/h)^2$ is large at small thickness, the deformation is pure bending almost everywhere and there is non-zero gaussian curvature only at singularities where the energetically expensive stretching is localized.

Here we study a plate initially curved in one direction (the x -direction), being clamped at two of its extremities (Fig. 1). The clamping provides two tunable parameters: the distance between the clamped sides d ; and the two equal angles (α) between the sheet and the horizontal direction at $x = \pm d/2$. The other two sides of the plates, at $y = \pm L/2$, are free. The plate is cylindrical (Fig. 1a). The cylinder axis is in the y -direction. A conical tip then pushes down the plate at its centre. The control parameter is the vertical displacement, Z , of the centre. This problem can be considered as the two-dimensional extension of the elastic arch investigated by Pippard^{12,13}. However, whereas the elastic arch has only bending deformations, the plate here may have stretching deformations as well.

Indeed, when the plate centre is displaced, two d-cones linked by an inverted ridge appear (Fig. 1a, e). As the vertical displacement of the centre Z is increased, the two d-cones move towards the clamped boundaries. The plate is still invariant by the x and y mirror symmetries. When the vertical displacement reaches about $Z = 10$ mm, a continuous transition occurs, breaking the two mirror symmetries. Two new d-cones appear, so that the four d-cones form a diamond which rotates about the centre (Fig. 1b, f). This rotation is either to the right or to the left. The plate is now

invariant by the 0-symmetry $(x, y) \rightarrow (-x, -y)$. At about $Z = 11.5$ mm a second continuous transition breaks all symmetries. The 4 d-cones move so that at $Z = 12$ mm they form a trapezoid (Fig. 1c, g). The trapezoid's larger edge is either to the left or to the right. The mirror symmetry $y \rightarrow -y$ is recovered. The trapezoid grows until two of its vertices reach the free sides of the sheet ($Z = 13$ mm). Then there is a discontinuous transition such that the plate becomes cylindrical (Fig. 1d). Note that these patterns are observed when the pushing tip is exactly at the centre of the plate (with a tolerance of about 0.1 mm). Once the adjustment is made, different realizations of the experiment (under the same conditions) show the two possible states following each continuous transition. The situation described here is the more general one: if the plate is long enough, all regimes are observed. If the length of the plate is decreased, first the trapezoid and then the diamond regime disappear: the d-cones reach the free boundaries and the discontinuous transition to a cylinder occurs earlier.

In elasticity theory, large deformation of plates are usually described by the Föppl von Kármán equations (FvK)¹⁴. These nonlinear equations are notoriously difficult, mainly because they

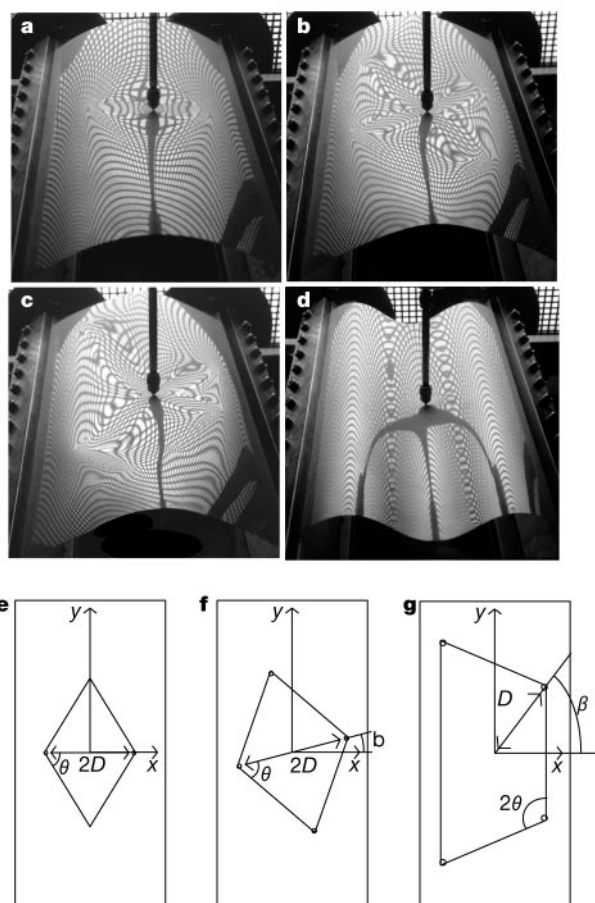


Figure 1 Observed patterns with a Mylar sheet of thickness $h = 0.35$ mm, Young's modulus $E = 3.8 \times 10^9$ N m⁻² and Poisson ratio $\nu = 0.4$. The aspect ratio is 2 (length $L = 35$ cm and width $W = 17.5$ cm). $\alpha = 20^\circ$ and $d = 16.5$ cm (see text for definitions). The results depend only slightly on the dimensions of the sheet. **a**, Vertical displacement of the centre $Z = 4$ mm: two d-cones located at $(x, y) = (\pm D, 0)$. **b**, $Z = 11.5$ mm: four d-cones forming a diamond at large rotation angle ($\beta \sim 30^\circ$). Two ridges link the d-cones. **c**, $Z = 12.5$ mm: 4 d-cones forming a trapezoid. **d**, $Z = 15$ mm: cylindrical shape. As Mylar has a large elastic limit there are no plastic (irreversible) deformations during this evolution. The experiment can thus be reproduced with the same sheet. **e–g**, Positions of the d-cones (black dots) in the **a**, **b** and **c**, respectively. The schemes give the definition of D , β and θ in the three regimes.

involve high-order derivatives and two types of deformations (stretching and bending) with energies of different orders of magnitude. A successful numerical procedure to overcome these difficulties has been proposed¹⁵. The plate is represented by finite elements and its energy is minimized. To obtain numerical convergence, the variables are changed by diagonalization of the energy hessian (re-conditioning) two or three times during the procedure.

Here, we compare experimental results and numerical solutions of the FvK equations. In Fig. 2, we show two numerical solutions and their energy as Z increases. Stretching is confined to the centre of the plate and to the d-cone tips. Bending dominates elsewhere. The pushing tip gives a spherical shape locally (non-zero Gauss curvature) to the plate so that there is stretching at the centre.

The experimental and the theoretical force versus displacement curve are represented in Fig. 3. The slope in zone II is larger than in zone I. This is explained by the numerics: some bending energy is localized near the boundaries. The agreement between numerics and experiment is good until zone III. So, we first concentrate on regimes I and II.

The contributions of the singularities dominate the energy of the plate. The energy of a ridge⁴ and the energy of a d-cone^{5–8} are given respectively by

$$E_r = F_r \kappa (R/h)^{1/3} \phi^{7/3} \quad (1)$$

$$E_d = F_d \kappa \ln(R/R_c) \phi^2 \quad (2)$$

Here $\kappa = Eh^3/12(1 - \nu^2)$ is the bending modulus, R the size of the singularity, ϕ its strength (the complement of the angle between the two sides of the ridge and the complement of the tip angle of the cone), $R_c = (\kappa/Eh)^{1/6} R^{2/3} \phi^{-1/3}$ is the radius of the core of the d-cone, $F_r \approx 1$ and $F_d \approx 100$ are shape factors. Under our conditions, where $h/R \approx 10^{-3}$ and $\phi \approx 0.1$, the energy of the ridges is negligible. As a consequence, the energy of the plate will be given by the energy of the d-cones.

Now we show that singularities are formed as soon as $Z \approx h$. We estimate the transition from the linear regime (the regime where the linearization of the FvK equations holds¹⁴) to the regime in which

two d-cones appear. If $S = r^2$ is the area of the deformed zone, the strain and the curvature are given by $s \approx Z^2/r^2$ and $c \approx Z/r^2$. So the stretching and the bending energy scale as $E_s \approx Ehs^2S \approx EhZ^4/r^2$ and $E_b \approx Eh^3c^2S \approx Eh^3Z^2/r^2 \approx Eh^3Z^2/r^2$ respectively. The comparison between these two expressions shows that the linear regime is only valid when $Z^2 < h^2$. The regime without singularities should correspond to the small region $Z \leq 0.35$ mm in Fig. 3.

In the two-d-cones regime, the central zone of the plate is diamond-shaped with two d-cones at the vertices where $\gamma = 0$. The angular opening (Fig. 1e) of each d-cone (the angular opening of the inverted curvature zone) is always in the range $2\theta = 100$ – 120° , the same as for a single d-cone^{7,8}. To find the positions of the d-cones, we use the following geometrical argument. Consider a section of the undeformed plate by the plane $\gamma = 0$. Approximate it by its osculating parabola $z = kx^2$. When the centre is pushed, the arc $x \in (-D + \Delta, D - \Delta)$ of the parabola is transformed into the ridge of length $2D$ linking the two d-cones. The equality of the arc length and the ridge length yields $\Delta \approx 2/3k^3D^3$, so that

$$kD \approx \sqrt{kZ}(1 + 2/3kZ) \quad (3)$$

Here Z is the imposed vertical displacement. This equation is checked experimentally and numerically in Fig. 4, so that the geometrical description holds. In the x -section, the angle between the generatrices of a d-cone is $\psi \approx 2kD$. Now we must find the size R of the d-cone. Near its tip, the curvature due to the d-cone ψ/r is large (r is the distance to the tip). Far from the tip, the initial curvature of the plate $2k$ dominates. R is given by the matching $\psi/R \approx 2k$. Here the geometry gives all the characteristics of the d-cones; we only need to compute the energy from equation (2). If we use the solution found for the single d-cone^{6–8}, $\phi = \psi/\mu$, $\mu = 4.8$ and $F_d = 67$. The solid line of Fig. 3 represents the force $F = dE_d/dZ$ calculated to the lowest order in kZ ($k = 5.7 \text{ m}^{-1}$). This description is only valid while the radius R of the d-cone is smaller than its distance to the boundary. This condition $D + R = W/2$ gives $Z_c = 10.9$ mm.

The rotation of the d-cones, leading to the diamond regime, can now be explained by a simple geometric argument. Consider the two d-cones with aperture angle 2θ and a section of the plate by the plane $\gamma = \tan\beta x$ (Fig. 1g). When $Z = 0$, this section is approximated by $z = k\cos^2\beta s^2$, where s is the coordinate along the section. The geometrical results obtained for the 2 d-cones regime (equation

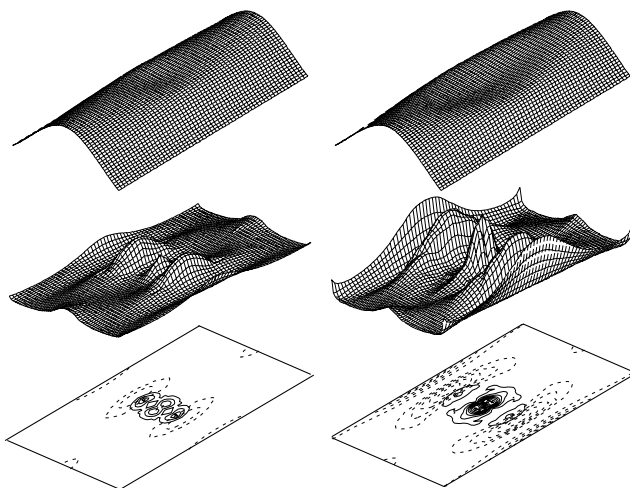


Figure 2 Two numerical situations with $\alpha = 20^\circ$ and $d = 16.5$ cm. Left, $Z = 5$ mm; right, $Z = 12$ mm. The procedure is a finite-element minimization of the plate energy, using reconditioning to obtain convergence. 471 minimization variables were used. Top, numerical configurations of the plate. Middle, energy density. A maximum at the centre (at the pushing tip) between two lateral maxima corresponding to two d-cones. For $Z = 12$ mm, some energy is localized at the clamped boundaries $(x, y) = (\pm d/2, 0)$. Bottom, contour lines of stretching energy (solid lines) and bending energy (dashed lines). The bending energy dominates except on very confined regions at the centre of the plate and the tips of the d-cones.

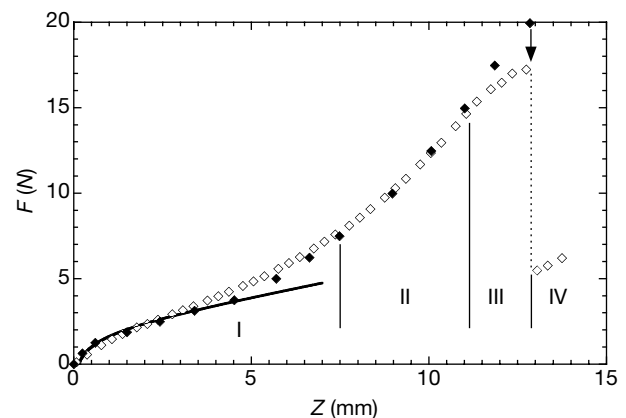


Figure 3 Force F versus displacement of the centre Z with $\alpha = 20^\circ$ and $d = 16.5$ cm. Force is measured with a piezoelectric cell allowing 10^{-2} N precision. Open diamonds, experimental result; filled diamonds, the numerics. The black arrow shows the numerical transition to regime IV. Solid line, theoretical prediction from equations (2) and (3), $F = 0.65 \ln(10.3Z)$. Regime I, two d-cones are observed. Regime II, almost linear, with a larger slope than in I. Regime III, the numerics overestimate the energy. Regime IV, cylindrical situation with no singularities.

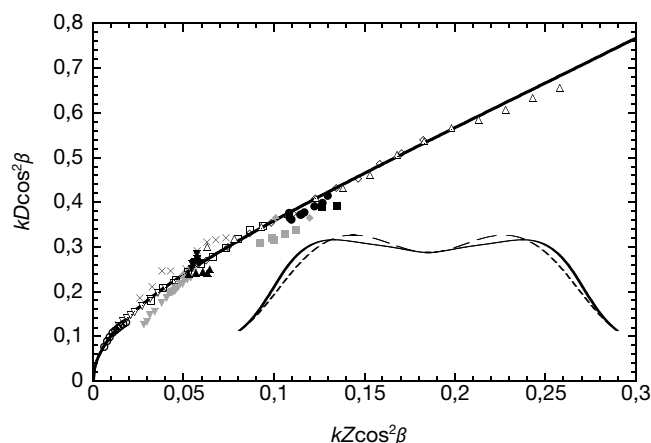


Figure 4 Distance between the dominant d-cones and the centre with various α and d values. k is half the mean curvature at the plate centre when $Z = 0$. Crosses, numerics; the other symbols are experimental. Open symbols, two-cones regime. Grey symbols, diamond regime. Black symbols, trapezoid regime. Solid line, theoretical prediction from equation (3). Inset curve ($Z = 6$ mm), x cross-section of the plate (solid line); elastic arch (dashed line).

(3)) hold when k is replaced by $k \cos^2 \beta$ as is checked in Fig. 4. To lowest order in kZ , the d-cone radius is $R \approx D \approx \cos \beta \sqrt{Z/k}$. If the interaction between the d-cone and the clamped boundary is strong, the d-cone is constrained to move to a new position $\beta \neq 0$, in which the periphery of the d-cone (a circle of radius R) exactly touches the boundary of the plate, $D \cos \beta + R = W/2$. For Z near to Z_c , this condition leads to

$$\beta \approx \pm 4\sqrt{2/3/kW} \sqrt{k(Z - Z_c)} \quad (4)$$

Here $4\sqrt{2/3/kW} = 3.3$ and $Z_c = 10.9$ mm compare well with the experimental $\beta = 3.1\sqrt{k(Z - Z_c)}$ with $Z_c = 10$ mm.

Now consider the second pair of d-cones. When their peripheries touch the clamped boundaries, $\beta + \theta \approx \pi/2$ so that $Z = Z_i \approx k(W/2/\sin \theta / (1 + \sin \theta))^2 = 15$ mm (with $\theta = 63^\circ$) which is comparable to the experimental $Z_i = 12$ mm. For $Z \geq Z_i$ all d-cones have to move along the clamped boundaries. If two of them are constrained to have an angular opening of 2θ , the only possible shape is a trapezoid. Compared to the two-d-cones regime, energy is gained in the diamond, and then the trapezoid regime. These configurations were not observed numerically, probably owing to the discretization. For this reason the numerics overestimates the force in region IV of Fig. 3. At $Z \geq 13$ mm the plate becomes cylindrical (pure bending deformations) so that the force jumps to a much smaller value (Fig. 3). This transition is well predicted by the numerics. The energy of the plate is then the same as the elastic arch¹³.

The experiment we have performed is typically a controlled version of what happens when a car is bumped. The difference is in the material; in metals the elastic limit is easily exceeded at the d-cones and then at the ridges. The first deformation of a plate is either cylindrical or conical. Most further deformations lead to the formation of pairs of d-cones and to the diamond-shaped deformation we have studied here. The observed shapes are determined by the geometry, and the energy of the plate can be computed using only the energy of the singularities. □

Received 22 February; accepted 24 July 2000.

1. Pogorelov, A. V. *Bendings of Surfaces and Stability of Shells*. (Translations of Mathematical Monographs No. 72, American Mathematical Society, 1988).
2. Witten, T. A. & Li, H. Asymptotic shapes of a fullerene ball. *Europhys. Lett.* **23**, 51–55 (1993).
3. Lobkovsky, A. E., Gentges, S., Li, H., Morse, D. & Witten, T. A. Scaling properties of stretching ridges in a crumpled elastic sheet. *Science* **270**, 1482–1485 (1995).
4. Lobkovsky, A. E. Boundary layer analysis of the ridge singularity in a thin plate. *Phys. Rev. E* **53**, 3750–3759 (1996).

5. Ben Amar, M. & Pomeau, Y. Crumpled paper. *Proc. R. Soc. Lond. A* **453**, 729–755 (1997).
6. Chaieb, S. & Melo, F. Experimental study of developable cones. *Phys. Rev. Lett.* **80**, 2354–2357 (1998).
7. Cerda, E. & Mahadevan, L. Conical surfaces and crescent singularities in crumpled sheets. *Phys. Rev. Lett.* **80**, 2358–2361 (1998).
8. Cerda, E., Chaieb, S., Melo, F. & Mahadevan, L. Conical dislocations in crumpling. *Nature* **401**, 46–49 (1999).
9. Chaieb, S. & Melo, F. Experimental study of crease formation in an axially compressed sheet. *Phys. Rev. E* **56**, 4736–4745 (1997).
10. Pomeau, Y. & Rica, S. Plaques très comprimées. *C.R. Acad. Sci. Paris* **325**, Série IIb, 181–187 (1997).
11. Pauchard, L. & Rica, S. Contact and compression of elastic spherical shells. *Phil. Mag.* **B 78**, 225–233 (1998).
12. Pippard, A. B. The elastic arch and its modes of instability. *Eur. J. Phys.* **11**, 359–365 (1990).
13. Patricio, P., Adda-Bedia, M. & Ben Amar, M. An elastica problem: Instabilities of an elastic arch. *Physica D* **124**, 181–187 (1998).
14. Landau, L. D. & Lifshitz, E. M. *Theory of Elasticity* (Pergamon, Oxford, 1986).
15. Patricio, P. & Krauth, W. Numerical solutions of the Von Karman equations for a thin plate. *Int. J. Mod. Phys. C* **8**, 427–434 (1997).

Acknowledgements

We thank L. Quartier and M. Haddad for their valuable help. This work was finished during the stay of M.B.A. and A.B. at the Department of Mechanical Engineering of the MIT; M.B.A. and A.B. are grateful for the hospitality they received.

Correspondence and requests for materials should be addressed to A.B. (e-mail: arezki.boudaoud@lps.ens.fr).

Interconversion of single and double helices formed from synthetic molecular strands

Volker Berl^{*†}, Ivan Huc^{‡*}, Richard G. Khoury^{*}, Michael J. Krische^{*§} & Jean-Marie Lehn^{*}

^{*} Laboratoire de Chimie Supramoléculaire, ISIS, Université Louis Pasteur, 4 rue Blaise Pascal, F-67000 Strasbourg, France

[†] Forschungszentrum Karlsruhe GmbH, Institut für Nanotechnologie, Postfach 3640, D-76021 Karlsruhe, Germany

[‡] Institut Européen de Chimie et Biologie, ENSCPB, Av. Pey Berland, F-33402 Talence Cedex, France

Synthetic single-helical conformations are quite common, but the formation of double helices based on recognition between the two constituent strands is relatively rare. Known examples include duplex formation through base-pair-specific hydrogen bonding and stacking, as found in nucleic acids and their analogues, and polypeptides composed of amino acids with alternating L and D configurations^{1,2}. Some synthetic polymers³ and self-assembled fibres⁴ have double-helical winding induced by van der Waals interactions. A third mode of non-covalent interaction, coordination of organic ligands to metal ions^{5–7}, can give rise to double, triple and quadruple helices, although in this case the assembly is driven by the coordination geometry of the metal and the structure of the ligands, rather than by direct inter-strand complementarity. Here we describe a family of oligomeric molecules with bent conformations, which exhibit dynamic exchange between single and double molecular helices in solution, through spiral sliding of the synthetic oligomer strands. The bent conformations leading to the helical shape of the molecules result from intramolecular hydrogen bonding within 2'-pyridyl-2-pyridinecarboxamide units^{8–12}, with extensive intermolecular

§ Present address: University of Texas at Austin, Department of Chemistry and Biochemistry, Austin, Texas 78712, USA.

aromatic stacking stabilizing the double-stranded helices that form through dimerization.

In the linear heptamers **1** to **3**, curvature of the strand should lead to a helical shape of the molecules (Fig. 1). Modelling studies suggest that coiling should extend to nearly one and a half turns. A 0.5-mM solution of **1** in CDCl₃ features a sharp ¹H NMR spectrum consistent with the presence of a single well defined species (Fig. 2a). The spectrum indicates that, on average, rings α, β and γ have the same environments as α', β' and γ' respectively. Comparison of the spectrum of **1** to that of its shorter analogue **4** shows that a number of signals are substantially shifted upfield, suggesting intramolecular stacking interactions in **1**. For example, the terminal amide signals CH₂CONH are found at 2.39 and 8.42 p.p.m. in **4**, and at 1.97 and 7.54 p.p.m. in **1**. Similarly, despite the almost symmetrical substitution of the β/β' rings, one of the two aromatic protons of these rings is shifted upfield by 0.5 p.p.m. from the other (7.40 and 7.88 p.p.m. compared to 7.92 p.p.m. for protons of the β ring of **4**).

These features are consistent with the single-helix structure of **1** shown in Fig. 1. Indeed X-ray diffraction analysis indicated that **3**, a less soluble analogue of **1**, crystallized from DMSO/CH₃CN as a single-helix entity (Fig. 3a). Conformationally stabilized single molecular helices presenting up to four turns have been described recently^{13–16}.

Upon concentrating a solution of **1** in chloroform, a second set of signals appears on the ¹H NMR spectrum (Fig. 2b–d), indicating the presence of another species in slow exchange on the NMR timescale with the single-helix entity. Heating an 8.2-mM solution causes an increase of the proportion of the latter, up to over 98% at 55 °C, without any coalescence. The concentration and temperature dependence of the proportions between the two species suggests

that the new signals correspond to an aggregate formed by two or more helical monomers. The exchange is clearly visible on the ¹H nuclear Overhauser enhancement spectroscopy (NOESY) spectra, which display cross-peaks between each signal of the monomer and each signal of the aggregate. Saturation transfer experiments yielded an exchange rate of 8.7 s⁻¹ at 25 °C.

In the NMR spectra, the protons of the aggregate are overall more strongly shielded than the protons in monomeric species, suggesting that intermolecular π–π aromatic stacking is important in the aggregation process as expected for compounds containing several aromatic residues. However, simple stacking of independent helices should be fast on the NMR timescale and lead to a mixture of various oligomers instead of a well defined aggregate. Instead, aggregation apparently requires considerable conformational changes in the molecules. For these reasons, we proposed the formation of a double helix consisting of two intertwined monomeric strands, which would stack all along their length, so that association and dissociation would require winding and unwinding of the helical monomeric strands. A related process has been described for the formation of the double-helix gramicidin A dimer^{2,17}.

The proportions between the monomer and its aggregate at different concentrations (1–25 mM; see Fig. 2) are consistent with a dimerization constant *K*_{dim} of 25–30 l mol⁻¹ in CDCl₃ at 25 °C. This value is strongly solvent dependent, being about 300, 22 and 95 l mol⁻¹ for CD₂Cl₂, 1/9 CDCl₃/CCl₄, and 3/7 CDCl₃/C₆D₁₂ respectively. In all these solvents, however, increasing the concentration leads to broadening of the signals or precipitation, and not to analysable spectra of the aggregates.

The NMR spectra of the less soluble analogue **3** feature similar concentration dependence and slow exchange between monomeric and dimeric species (not shown). The proportions between these signals are in agreement with a dimerization constant of 110–120 l mol⁻¹ in CDCl₃ at 25 °C. That these processes observed in solution do correspond to a double helix formation is supported by a crystal structure of the same compound **3**, which crystallized from nitrobenzene/heptane as a double helix (Fig. 3b).

As expected, the double-helix structure allows considerable overlap between the aromatic groups of each monomeric helical strand, with an average π–π stacking distance of 3.5 Å, corresponding to

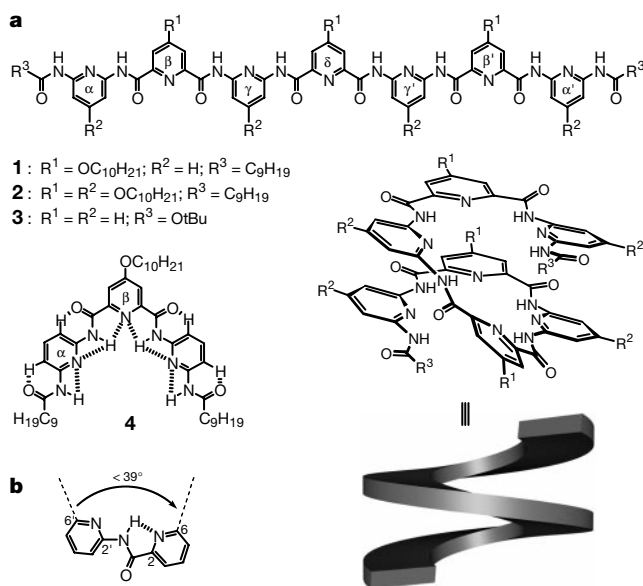


Figure 1 Structures and folding of oligopyridinecarboxamides. **a**, Formula of various oligomers and helical shape of a heptamer. These compounds were synthesized according to the scheme used for isophthalate analogues²⁰. The decyloxy chains ensure the excellent solubility of **1** and **2** in chlorinated or aromatic solvents, and even in pure alkanes for **2**. **b**, Conformation of the 2'-pyridyl-2-pyridinecarboxamide motif. Conjugation of the aromatic rings with the amide moiety and intramolecular hydrogen bonding of the amide hydrogen to one pyridine ring is expected to stabilize one conformer in which substituents of the pyridine rings at the 6 and 6' positions protrude on the same side of the molecule. Modelling studies indicate that these substituents define an angle reduced to 39°. In the solid state, this angle can be as low as 35° (ref. 8).

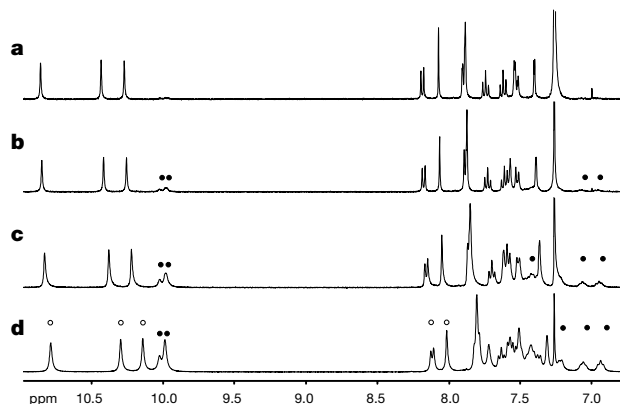


Figure 2 400-MHz ¹H spectra of CDCl₃ solutions of **1** at various concentrations at 25 °C. Some of the signals assigned to the monomer (unfilled circles), and to the dimer (filled circles) are labelled. **a**, 0.91 mM. The formation of intramolecular hydrogen bonds results in a strong deshielding of three of the four different amide protons (10.86, 10.43 and 10.27 p.p.m.), the terminal amide hydrogen being unable to undergo similar interactions; **b**, 2.7 mM; **c**, 8.2 mM; **d**, 24.5 mM. Nuclear Overhauser enhancement spectroscopy (NOESY) experiments confirm that the triplets observed at 7.06 and 6.94 p.p.m. correspond to the signals of the protons in position 4 of the α/α' and γ/γ' rings in the dimer.

van der Waals contact. The coulombic and van der Waals forces associated with aromatic stacking thus seem to promote interstrand attractive interactions. Most hydrogen bonds occur within the same strand and stabilize the helical shape of each monomer. Four direct interstrand $\text{NH}\cdots\text{N}$ hydrogen bonds occur at the ends of the duplex (3.19–3.32 Å and 3.21 Å), along with two bridging $\text{NH}\cdots\text{O}$ hydrogen bonds (3.01 and 3.09 Å) to a water molecule bound to the polar inner rim of the duplex. This structure contrasts with the structure of DNA, in which hydrogen bonding determines interstrand association, and stacking takes place mainly within each of the two strands.

The presence of bound water in the crystal lattice is consistent with previous crystallographic data^{8,11}, and with ^1H NMR solution studies of **1**, which feature a broad singlet between 3 and 5 p.p.m. This water molecule solvates the polar functions of **1**, and may contribute to the helical pre-organization. However, molecular modelling studies suggest that the water molecule is not necessary for helicity induction. Furthermore, the ratio of dimer versus monomer of **1** was found to increase significantly upon partial removal of water from the solvent (CDCl_3 or C_6D_6), whereas increasing the amounts of water led to the dissociation of the duplexes. This is possibly owing to the destabilization of the helical strand conformation as a result of water competing with intramolecular hydrogen bonding within **1**.

It should be emphasized that these solid-state characterizations of single and double molecular helices are from the same

compound **3**, crystallized from different media. The single helix crystallized from a rather polar solvent mixture, whereas the double helix crystallized from a less polar solvent mixture. This is consistent with a polarity-dependent equilibrium and supports the idea that these helices are indeed the two equilibrating species observed by NMR and that their formation involves polar interactions.

Compound **2**, bearing four additional decyloxy chains, was expected to be more soluble in non-polar solvents. Its ^1H NMR spectra in CDCl_3 suggests that it also undergoes dynamic equilibrium between a single-helix monomer and a double-helix dimer (see Supplementary Information). In this case however, the dimer remains the major species at concentrations as low as 300 μM . Quantitative analysis of the NMR data led to a dimerization constant $K_{\text{dim}} = 6.5 \times 10^4 \text{ l mol}^{-1}$, which is three orders of magnitude larger than that of **1** in the same solvent. In toluene- D_8 , CD_2Cl_2 and $\text{C}_2\text{D}_2\text{Cl}_4$, this value was found to be $5.5 \times 10^4 \text{ l mol}^{-1}$, $1.0 \times 10^5 \text{ l mol}^{-1}$, and $1.6 \times 10^5 \text{ l mol}^{-1}$ respectively. This marked effect may result from a larger number of interactions between the side chains as well as an increase in the interactions between

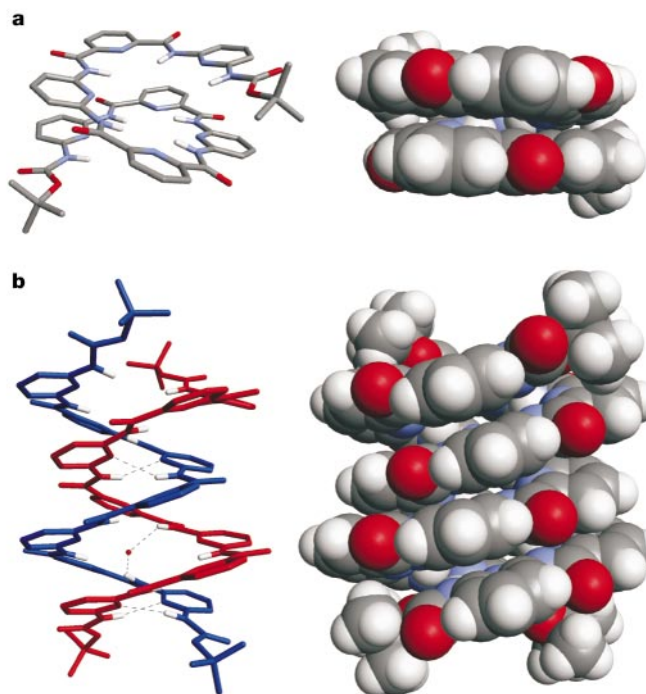


Figure 3 Crystal structures of the single helix and the double-helix dimer of **3**. The structures were determined by X-ray diffraction of two independent crystals. Solvent molecules included in the crystals and CH hydrogens in the stick representations are omitted for clarity. Carbon, grey; hydrogen, white; nitrogen, blue; oxygen, red. **a**, Single helix. Winding of the strand reaches one and a half turns, where one turn is generated by approximately five pyridine-amide units. Intramolecular $\pi\cdots\pi$ stacking is possible over half a helical turn, which is consistent with the upfield shift of some NMR signals observed in solution. **b**, Double-helix dimer. The stick representation with colour-coded strands shows interstrand hydrogen bonds (black dotted lines). Winding of the two strands results in a nearly two-turn duplex, with 16.7 Å in length, in which only four pyridine-amide units are needed per turn. Single/double helix interconversion corresponds to a spring-like compression/extension motion of each strand.

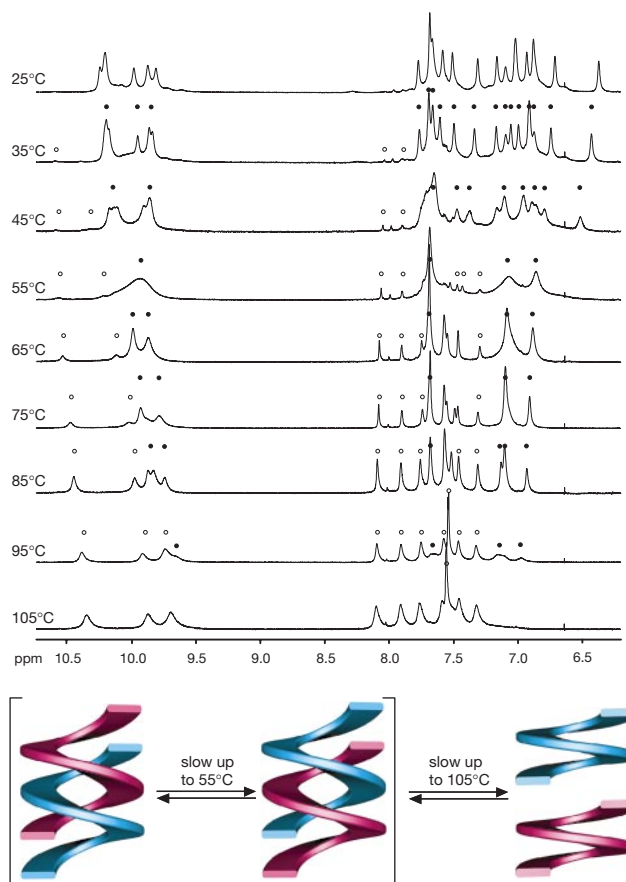


Figure 4 Formation and interconversion of double helices of **2**. Top, 400-MHz ^1H NMR spectra of **2** (8.2 mM in $\text{C}_2\text{D}_2\text{Cl}_4$) at various temperatures. Some of the signals are assigned to the dimer (filled circles), to the monomer (unfilled circles), and to a minor impurity (asterisk). Bottom, schematic representations of the interconversion between two identical forms of dissymmetrical double helices by a sliding motion, and their dissociation into two single helices. In the dissymmetrical double-helical structures depicted, rings α and α' , β and β' , and γ and γ' of an individual strand are situated in different local environments, and thus differ in their ^1H NMR signals. Their interconversion becomes rapid on the NMR timescale at 55°C, and the signals associated with analogous protons in rings α and α' , β and β' , and γ and γ' coalesce. The dissociation of the double helix into single helices is slow on the NMR timescale up to 105°C.

aromatic rings due to the electron donor character of the decyloxy substituents in **2**.

We note two important differences between the proton signals of (**2**)₂ and (**1**)₂ in CDCl₃. The first refers to the broadness of the signals of (**2**)₂. This cannot be the result of exchange with monomer **2** whose signals would then broaden as well. More probably, it reflects slow conformational changes within (**2**)₂ or equilibration with other larger aggregates. The second difference is the non-equivalence of α/α' , β/β' , and γ/γ' rings in dimer (**2**)₂, which is apparent in CDCl₃ and clearly confirmed in C₂D₂Cl₄. In this latter solvent, the dimer signals are sharp at 25 °C and clearly indicate a dissymmetrical structure. In the hydrogen-bonded amide region (9.5–10.5 p.p.m.), six distinct signals of equal intensity can be counted instead of three for a symmetrical structure (see Supplementary Information). Between 6.5 and 8.5 p.p.m., the spectrum shows 16 partly overlapping singlets, corresponding to 14 different aromatic protons, and two different terminal amides NH. When we varied the concentration, the proportions between these signals remain unchanged. NOESY experiments show cross-peaks indicating slow exchange of half of the signals with the other half.

The evolution of the spectrum of **2** with temperature is shown in Fig. 4. At 25–35 °C the spectrum consists essentially of sharp signals of a dissymmetrical dimer. Above 45 °C, as the monomer proportion increases, its signals start to be clearly visible. Simultaneously, the dimer signals broaden substantially to coalesce at 55 °C. This spectrum is reminiscent of that observed in CDCl₃ at room temperature. Above 55 °C, the dimer signals sharpen again but their number is decreased. Although several signals overlap, the spectrum seems consistent with an averaged symmetrical structure such as the one observed for compound **1** (Fig. 2). At 95 °C, the monomer is by far the major species, and is still in slow exchange with the dimer. At 105 °C, the dimer peaks have disappeared, owing to their very low intensity, or because coalescence with the monomer has finally been reached.

In view of these data, the dimerization of compounds **1** and **2** appears to give rise to dissymmetrical double helices, so that the two ends of an individual strand are no longer equivalent (Fig. 4). Sliding of the monomers along one another in a spiralling motion and without dissociation allows each of the two ends of an individual strand to switch between states or local environments. Above a certain temperature, these equilibria become rapid on the NMR timescale. A related sliding process has been identified in the double-helix gramicidin A dimer^{18,19}. We also note that such an interconversion would amount to the rotational translation of one strand with respect to the other, and thus it constitutes a molecular mechanical motion. □

Methods

Crystallographic data

Single crystals of **3** (single helix), [C₅₁H₄₇H₁₅O₁₀·2(C₂H₆SO)·2(H₂O)·1.5(C₂H₃N)], were grown from acetonitrile/dimethylsulphoxide. Crystals were placed in oil and a single colourless crystal of dimensions 0.25 × 0.16 × 0.15 mm was selected, mounted on a glass fibre, and placed in a low-temperature N₂ stream. The unit cell was monoclinic with a space group of C2/c. Cell dimensions: $a = 33.952(7)$ Å, $b = 13.510(3)$ Å, $c = 27.970(6)$ Å, $\alpha = \gamma = 90^\circ$, $\beta = 97.75(3)^\circ$, $V = 12713(4)$ Å³ and $Z = 8$ (FW is 1307.91, $\rho = 1.367$ g cm⁻³). Reflections were collected from $1.0^\circ \leq \theta \leq 27.6^\circ$ for a total of 14,359 of which 7,699 were unique ($R_{\text{int}} = 0.089$) having $I > 4\sigma(I)$; number of parameters is 802. Final R factors were $R_1 = 0.117$ (based on observed data, $wR_2 = 0.254$ (based on all data), GOF = 1.067, maximal residual electron density is 0.933 e Å⁻³). Single crystals of **3** (double helix), [C₅₁H₄₇H₁₅O₁₀·3(C₆H₅NO₂)·(H₂O)], were grown from heptane/nitrobenzene. The crystals were placed in oil and a single colourless crystal of dimensions 0.18 × 0.15 × 0.15 mm was selected, mounted on a glass fibre, and placed in a low-temperature N₂ stream. The unit cell was triclinic with a space group of P1. Cell dimensions: $a = 18.980(4)$ Å, $b = 19.230(4)$ Å, $c = 19.980(4)$ Å, $\alpha = 81.52(3)^\circ$, $\beta = 75.11(3)^\circ$, $\gamma = 81.59(3)^\circ$, $V = 6926(2)$ Å³ and $Z = 4$ (FW is 1417.38, $\rho = 1.350$ g cm⁻³). Reflections were collected from $1.0^\circ \leq \theta \leq 27.01^\circ$ for a total of 30,136 of which 14,924 were unique ($R_{\text{int}} = 0.035$) having $I > 4\sigma(I)$; number of parameters = 1865. Final R factors were $R_1 = 0.099$ (based on observed data, $wR_2 = 0.214$ (based on all data), GOF is 1.005, maximal residual electron density is 0.600 e Å⁻³).

X-ray diffraction data for **3** (single and double helix) was collected on a Nonius Kappa

charge-coupled device (CCD) diffractometer with a graphite monochromatized MoK α radiation ($\lambda = 0.71071$ Å). φ scans, at 173 K. The structure solution of **3** (single and double helix) was determined using direct methods and refined (based on F^2 using all independent data) by full matrix least square methods (SHELXTL 97). Hydrogen atoms were included at calculated positions by using a riding model.

The crystal lattice of the single helix contains eight helical strands of **3** (four enantiomeric pairs) along with 16 molecules of DMSO, 12 molecules of acetonitrile and 16 molecules of water, out of which eight are located in the polar inner rim of the helices. The helical pitch of the helix is 3.5 Å and the pyridine–pyridine torsional angles are between 1.3° and 16.5° (average 10.1°).

The crystal lattice of the double-helix dimer contains four helical strands of **3** (two enantiomeric double-helix pairs), 12 molecules of nitrobenzene and two molecules of water bound to the polar inner rim of the double helix. Although the two strands of a duplex are crystallographically not equivalent, they are almost superimposable upon 180° rotation about the helical axis. The double-helix pitch is 7.2 Å and the pyridine torsional angles are opened up to between 17.9 and 34° (average 25.4°).

Received 11 April; accepted 17 August 2000.

- Di Blasio, B., Benedetti, E., Pavone, V. & Pedone, C. Regularly alternating L,D-peptides. The double-stranded right-handed antiparallel β -helix in the structure of *t*-Boc-(L-Phe-D-Phe)₄-OMe. *Biopolymers* **28**, 203–214 (1989).
- Langs, D. A. Three-dimensional structure at 0.86 Å of the uncomplexed form of the transmembrane ion channel peptide gramicidin A. *Science* **241**, 188–191 (1988).
- Kusanagi, H. X-ray and energy calculation studies on the packing-mode of double-stranded helices of isotactic poly(methyl methacrylate). *Polymer J.* **28**, 708–711 (1996).
- Engelkamp, H., Middelbeek, S. & Nolte, R. J. M. Self-assembly of disk-shaped molecules to coil-coil aggregates with tunable helicity. *Science* **284**, 785–788 (1999).
- Koert, M., Harding, M. & Lehn, J.-M. DNH Deoxyribonucleohelicates—self-assembly of oligonucleosidic double-helical metal complexes. *Nature* **346**, 339–342 (1990).
- Hasenknopf, B., Lehn, J.-M., Kneisel, B. O., Baum, G. & Fenske, D. Self-assembly of a circular double helicate. *Angew. Chem. Int. Edn Engl.* **35**, 1838–1840 (1996).
- Bell, T. W. & Jousselin, H. Self-assembly of a double-helical complex of sodium. *Nature* **367**, 441–444 (1994).
- Redmore, S. M., Rickard, C. E. F., Webb, S. J. & Wright, L. J. Ruthenium complexes of a simple tridentate ligand bearing two 'distal' pyridine bases. *Inorg. Chem.* **36**, 4743–4748 (1997).
- Hamuro, Y., Geib, S. J. & Hamilton, A. D. Novel folding patterns in a family of oligoantranilamides: Non-peptide oligomers that form extended helical secondary structures. *J. Am. Chem. Soc.* **119**, 10587–10593 (1997).
- Hamuro, Y., Geib, S. J. & Hamilton, A. D. Novel molecular scaffolds—formation of helical secondary structure in a family of oligoantranilamides. *Angew. Chem. Int. Edn Engl.* **33**, 446–448 (1994).
- Yu, Q. *et al.* Hydrogen bonded antiparallel beta-strand motifs promoted by 2,6-bis(carbamoyl)pyridine. *Chem. Commun.* 1467–1468 (1999).
- Mazik, M., Bläser, D. & Boese R. Hydrogen-bonding motifs in the crystals of secondary diamides with 2-amino-6-methyl and 2,6-diaminopyridine subunits. *Tetrahedron* **55**, 12771–12782 (1999).
- Bassani, D., Lehn, J.-M., Baum, G. & Fenske, D. Designed self-generation of an extended helical structure from an achiral polyheterocyclic strand. *Angew. Chem. Int. Edn Engl.* **36**, 1845–1847 (1997).
- Ohkita, M., Lehn, J.-M., Baum, G. & Fenske, D. Helicity coding: Programmed molecular self-organization of achiral nonbiological strands into multimeric helical superstructures: synthesis and characterization of alternating pyridine–pyrimidine oligomers. *Chem. Eur. J.* **5**, 3471–3481 (1999).
- Cuccia, L. A., Lehn, J.-M., Homo, J.-C. & Schmutz, M. Encoded helical self-organization and self-assembly into helical fibers of an oligoheterocyclic pyridine–pyridazine molecular strand. *Angew. Chem. Int. Edn Engl.* **37**, 233–237 (2000).
- Nelson, J. C., Saven, J. G., Moore, J. S. & Wolynes, P. G. Solvophobically driven folding of nonbiological oligomers. *Science* **277**, 1793–1796 (1997).
- Veatch, W. R. & Blout, E. R. The aggregation of gramicidin A in solution. *Biochemistry* **13**, 5257–5264 (1974).
- Sychev, S. V., Barsukov, I. I. & Ivanov, V. T. The double $\pi\pi$ 5.6 helix of gramicidin A predominates in unsaturated lipid membranes. *Eur. Biophys. J.* **22**, 279–288 (1993).
- Barsukov, I. I., Arseniev, A. S. & Bystrov, V. F. Spatial structures of gramicidin A in organic solvents. ¹H-NMR analysis of four species in ethanol. *Biorg. Khim.* **13**, 1501–1522 (1987).
- Berl, V., Krische, M. J., Huc, L., Lehn, J.-M. & Schmutz, M. Template-induced and molecular recognition-directed hierarchical generation of supramolecular assemblies from molecular strands. *Chem. Eur. J.* **6**, 1938–1946 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Fischer and A. DeCian for the use of the X-ray crystallographic facilities at ULP, Strasbourg. This work was supported by the CNRS, by a predoctoral fellowship (V.B.) from the Forschungszentrum Karlsruhe GmbH, Germany, and by a NSF-NATO postdoctoral fellowship (R.G.K.).

Correspondence and requests for materials should be addressed to I. H. (e-mail: ivan.huc@iecb-polytechnique.u-bordeaux.fr) or J.-M. L. (e-mail: lehn@chimie.u-strasbg.fr). Crystallographic data for **3** (excluding structure factors) have been deposited with the Cambridge Crystallographic Data Centre as supplementary publication no. CCDC-142810 (**3**, single helix) and CCDC-142811 (**3**, double helix). Copies of the data can be obtained free of charge on application to CCDC, 12 Union Road, Cambridge CB2 1EZ, UK.

Prototype systems for rechargeable magnesium batteries

D. Aurbach, Z. Lu, A. Schechter, Y. Gofer, H. Gizbar, R. Turgeman, Y. Cohen, M. Moshkovich & E. Levi

Department of Chemistry, Bar-Ilan University, Ramat Gan 52900, Israel

The thermodynamic properties of magnesium make it a natural choice for use as an anode material in rechargeable batteries, because it may provide a considerably higher energy density than the commonly used lead–acid and nickel–cadmium systems. Moreover, in contrast to lead and cadmium, magnesium is inexpensive, environmentally friendly and safe to handle. But the development of Mg batteries has been hindered by two problems. First, owing to the chemical activity of Mg, only solutions that neither donate nor accept protons are suitable as electrolytes; but most of these solutions allow the growth of passivating surface films, which inhibit any electrochemical reaction^{1–3}. Second, the choice of cathode materials has been limited by the difficulty of intercalating Mg ions in many hosts⁴. Following previous studies of the electrochemistry of Mg electrodes in various non-aqueous solutions^{1,5}, and of a variety of intercalation electrodes^{6,7}, we have now developed rechargeable Mg battery systems that show promise for applications. The systems comprise electrolyte solutions based on Mg organohaloaluminate salts, and Mg_xMo₃S₄ cathodes, into which Mg ions can be intercalated reversibly, and with relatively fast kinetics. We expect that further improvements in the energy density will make these batteries a viable alternative to existing systems.

Magnesium can be electrochemically deposited and dissolved reversibly in solutions of Grignard reagents (RMgX, where R is an alkyl or aryl group, and X is Cl or Br)^{8,9} in ether; this is because magnesium electrodes are not passivated in such systems⁵. However, as Grignard reagents are intrinsically strong reducing agents, they

cannot be used in batteries. In 1990, Gregory *et al.*¹⁰ reported attempts to construct rechargeable Mg battery systems based on Mg(BR₄)₂ solutions (where R can be various organic groups) and a Mg_xCoO_y intercalation cathode¹⁰. However, the results presented by these authors do not make a convincing case that these Mg–Mg_xCoO_y batteries are indeed practical, due to insufficient anodic stability of the electrolyte system used. Apart from Gregory's publication, we are not aware of any published research and development of integral, practical, rechargeable magnesium batteries.

We have developed a family of electrolyte solutions based on Mg organohaloaluminate salts—such as Mg(AlCl₃R)₂ and Mg(AlCl₂RR')₂, where R and R' are alkyl groups—in tetrahydrofuran (THF) or polyethers of the glyme family. The most effective alkyl groups in this context are C₄H₉ (butyl, Bu) and C₂H₅ (ethyl, Et) (see Methods). The room-temperature conductivity of these solutions at salt concentrations of approximately 0.3 to 0.5 M is high, in the range of several millisiemens, comparable to the electrolytes used in Li batteries¹¹.

Two characteristics of these solutions have emerged. First, the electrochemical deposition–dissolution of magnesium in these solutions is reversible at almost 100% efficiency. Second, the solutions exhibit an electrochemical window (the potential range in which the solution is electrochemically inactive) of more than 2.5 V, much wider than any electrolyte solution previously tested for rechargeable Mg batteries^{6–10}. Figure 1 compares the potentiodynamic behaviour of one of these electrolyte solutions, Mg(AlCl₂BuEt)₂ in tetrahydrofuran (THF), with the behaviour of two other types of electrolyte solutions in which Mg can be reversibly deposited: THF solutions of Mg(BPh₂Bu₂)₂ (ref. 10) and BuMgCl (ref. 1). These voltammograms indicate reversible Mg deposition–dissolution processes in all three solutions. However, the anodic stability of Mg(AlCl₂BuEt)₂ is clearly superior to the others. Hence, the incorporation of the electron-withdrawing halogen ligands on the organoaluminate core extends the electrochemical window of the solution without compromising the reversibility of the magnesium deposition–dissolution process.

We investigated the reversibility and efficiency of Mg anodes during repeated charge–discharge cycling in THF/Mg(AlCl₂BuEt)₂, using an electrochemical quartz crystal microbalance (EQCM). We modified this technique for use with highly reactive electrochemical systems¹². In these experiments, the electrochemical response (*I*, *V*, *t*) is measured together with mass accumulation and depletion on the electrodes at a resolution of a few nanograms per cm². Figure 2 shows typical results of these experiments (in constant-current

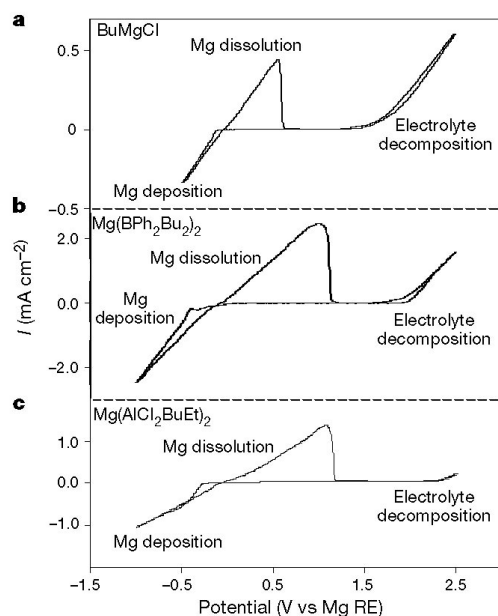


Figure 1 Comparison of cyclic voltammograms of electrolyte solutions in which magnesium can be deposited reversibly. A platinum working electrode, and Mg wire and strip reference and counter electrodes, were used. The potential scan rate was 5 mV s^{−1}. Solutions of the following in THF were used as electrolytes: **a**, BuMgCl 2 M; **b**, Mg(BPh₂Bu₂)₂ 0.25 M; **c**, Mg(AlCl₂BuEt)₂ 0.25 M. RE, reference electrode.

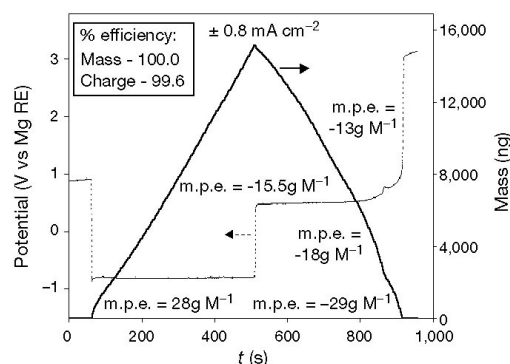


Figure 2 Results of typical galvanostatic EQCM experiments of Mg deposition–dissolution cycles with gold/quartz electrodes¹³. (The constant current was 0.8 mA cm^{−2}.) 0.25 M Mg(AlCl₂BuEt)₂ in THF was used as the electrolyte. Solid line (right ordinate): mass accumulated and depleted (mass versus time). Dotted line (left ordinate): potential profile (voltage versus time). The figure also shows mass accumulated and depleted (minus sign) per mole of electrons (m.p.e.) at different stages.

mode). From the mass accumulation and depletion, and the voltage profile versus time, both mass and charge balance are near zero, indicating practically 100% efficiency of the magnesium deposition–dissolution cycles. Figure 2 also indicates the mass per mole of electrons transferred (m.p.e.) at different stages of these processes. In the beginning of the deposition and at the end of dissolution, we find relatively high m.p.e. values due to a pronounced change in the surface roughness at these stages of the processes. Otherwise, the m.p.e. is approximately the equivalent weight of magnesium, 12 g per mole of electrons (as expected for Mg deposition–dissolution at nearly 100% efficiency).

The surfaces of magnesium films deposited from these electrolyte solutions were studied by *in situ* Fourier-transform infrared spectroscopy and *ex situ* by X-ray photoelectron spectroscopy. (Details of the experimental set-ups, as developed for studying highly reactive systems, are given in refs 13 and 14, respectively). Our results indicate that Mg deposits in these electrolyte systems do not develop stable surface films. We also studied magnesium deposition in these electrolytes by using *in situ* scanning electron tunnelling microscopy (STM), using a special experimental set-up¹⁵, and

obtained high-resolution images of magnesium deposits. Because successful use of STM requires electronic surface conductivity, the fact that we were able to study the morphology of Mg deposition processes in this way is additional evidence that the deposits do not develop passivating surface films.

We have also developed appropriate cathodes for these systems. The literature contains only a few practical studies on cathodes for rechargeable magnesium batteries^{4,10}. These cathodes suffer from either low charge capacity, extremely slow intercalation kinetics, or from potentials that are inaccessible in electrolytic solutions in which magnesium is reversibly deposited. Following comprehensive screening tests, we have found that Mo_3S_4 having a Chevrel phase structure^{16–18} is an excellent host material into which Mg ions can be reversibly intercalated electrochemically, at relatively fast kinetics, in a variety of non-aqueous Mg salt solutions (see Methods).

Figure 3 presents typical potentiodynamic behaviour of the $\text{Mg}_x\text{Mo}_3\text{S}_4$ electrodes in $\text{Mg}(\text{AlCl}_2\text{BuEt})_2$ solution in THF. The main figure shows the galvanostatic charge–discharge process, and a voltammogram is shown in the inset, together with an illustration of the structure of this compound based on X-ray

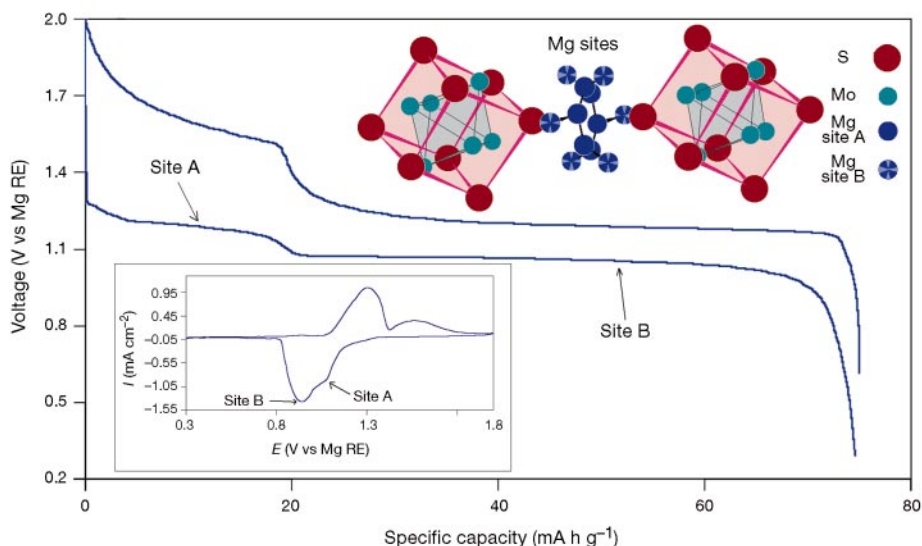


Figure 3 Typical electrochemical behaviour and the basic structure of the $\text{Mg}_x\text{Mo}_3\text{S}_4$ cathodes, $0 < x < 1$, corresponding to a maximal charge capacity of 122 mA h g^{-1} . The electrolyte was $0.25 \text{ M Mg}(\text{AlCl}_2\text{BuEt})_2$ in THF. A chronopotentiogram (main figure; voltage

versus capacity at constant current of 0.3 mA cm^{-2}) and a cyclic voltammogram (inset; 0.05 mV s^{-1}) of steady state Mg insertion–deinsertion cycles are shown.

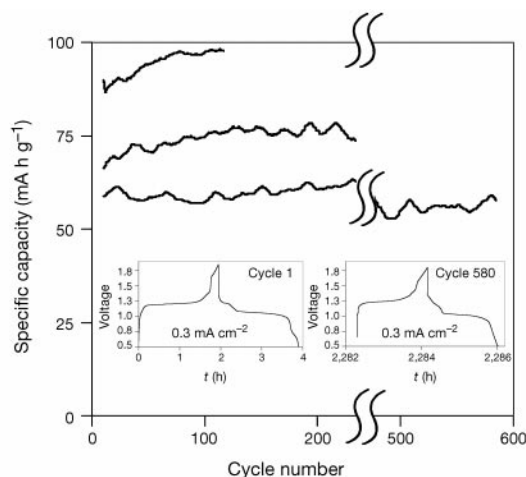


Figure 4 Performance of rechargeable $\text{Mg}-\text{Mg}_x\text{Mo}_3\text{S}_4$ coin-cell batteries during cycling at constant currents ($0.2-0.3 \text{ mA cm}^{-2}$). The electrolyte was $0.25 \text{ M Mg}(\text{AlCl}_2\text{BuEt})_2$ in THF. The specific capacity of the active mass is plotted against cycle number for three

cells, at different optimization levels, together with voltage profiles of representative charge–discharge cycles.

diffraction measurements and literature data^{16–18}.

As in other Chevrel compounds¹⁶, the crystal structure of the $\text{Mg}_x\text{Mo}_3\text{S}_4$ cathode may be considered as a stacking of Mo_6X_8 blocks. Each block contains an octahedral cluster of the molybdenum atoms inside a cube of sulphur atoms. Figure 3 shows that the sites for the guest Mg atoms are located in the channels between these blocks. An important feature of this structure is the delocalization of the sites of the guest atoms. There are 12 possible positions for Mg atoms between each set of two blocks, but only some of these sites may be occupied simultaneously due to geometrical and electrostatic limitations. In addition, guest atoms are weakly bound and move between these sites without the existence of an electric field, even at room temperature^{16–18}. Thus, the Chevrel phase has a stable structure for a wide range of intercalated metal content, with easy redistribution of electronic charge on each element (Mo or S) during insertion of cations. It also allows high mobility of the inserted metal ions, resulting in the feasibility of reversible Mg intercalation into $\text{Mg}_x\text{Mo}_3\text{S}_4$.

The process of magnesium insertion into this host material takes place in stages corresponding to the formation of new phases. As indicated in Fig. 3, these phase transitions are manifested by the electrochemical response of these processes as plateaus (chronopotentiometry) or peaks (cyclic voltammetry) on the electrochemical curves. A new intercalated phase differs from the previous one by the occupation of different sites in the host crystal with different insertion energy. By analogy to $\text{Li}_x\text{Mo}_3\text{S}_4$ (ref. 18), we assume that the first stage of intercalation (plateau at 1.2 V in Fig. 3) relates to the occupation of the so-called inner sites (sites A in Fig. 3) and the second stage (plateau at 1 V) relates to the outer ones (sites B in Fig. 3). Hence, the observed electrochemical behaviour of Mg insertion into Mo_3S_4 is well explained by the crystal structure of this material.

These developments pave the way for construction of complete rechargeable magnesium battery systems whose nominal voltage is between 1 and 1.3 V, and whose theoretical energy density is approximately 135 Wh kg^{-1} . We have assembled prototype rechargeable magnesium batteries in standard 'coin' cells (see Methods). The parameters explored included cathode load (5–100 mg active mass per cm^2), anode/cathode capacity ratio, rates, temperature range, and cycle life.

Figure 4 presents the typical charge–discharge behaviour of these batteries in terms of capacity versus cycle number, and the insets show two selected voltage profiles of representative, constant current charge–discharge cycles. The capacities shown reflect different levels of optimization. All three experiments represented in Fig. 4 were terminated for electrode analysis, and not because of electrode failure. In general, more than 2,000 charge–discharge cycles at 100% depth of discharge of the cathodes could be obtained at practical rates ($0.1\text{--}1 \text{ mA cm}^{-2}$) and over a wide temperature range ($-20^\circ\text{C} < T < 80^\circ\text{C}$), with capacity fading of less than 15%. The layer of electrolyte solution in the prototype batteries was very thin, and the ratio of electrode-area to solution-volume was similar to that found in any practical coin cell batteries or in batteries with a prismatic or spirally wound configuration. Hence, the stability observed over prolonged cycling shows the excellent compatibility of the various battery components.

These prototype Mg batteries have important advantages over other available rechargeable battery systems, in terms of safety and environmental considerations. They are constructed from abundant and environmentally friendly materials.

We note that these Mg batteries are not designed to compete with lithium batteries in terms of energy density for small-scale applications (for example, portable electronic equipment). However, Li batteries cannot replace lead–acid batteries for heavy load applications, due to safety problems and their relatively high cost. Taking into account the increasing needs of modern society for environmentally friendly, reliable, and cheap power sources in heavy load

applications, rechargeable 'green' Mg batteries that can deliver thousands of full charge–discharge cycles are the only battery systems to date that have a chance of rivaling the environmentally problematic lead–acid and nickel–cadmium batteries. The fraction of the active mass that is used in these Mg batteries is very high ($>90\%$ in the cathode and close to 100% in the anode). The percentage of passive additives needed in the cathode is very low (it may be reduced below 5%), and the amount of electrolyte solution is very small, as it does not participate in the battery reactions. We anticipate that, with good engineering, it should be possible to obtain large, rechargeable $\text{Mg/Mg}_x\text{Mo}_3\text{S}_4$ batteries with practical energy densities greater than 60 Wh kg^{-1} , which is considerably higher than that of the usual Ni–Cd and lead–acid batteries (about 40 Wh kg^{-1} ; ref. 19).

The present results are only the first step in the exploration of this technology. The energy density of rechargeable Mg batteries could be considerably increased by cathode modifications (for example, doping by electropositive atoms), which are expected to increase the voltage of these batteries. This work is in progress. □

Methods

The $\text{Mg}(\text{AlCl}_3\text{--}n\text{R}_n\text{R}')_2$ electrolytes were prepared by reacting MgR_2 (for example, MgBu_2) and AlCl_3R or AlCl_3 in hexane, followed by a complete evaporation of the hexane and dissolution in the selected ether solvent. The ether solvents are important in the stabilization of these salts. Furthermore, the electro-active cation is not simply solvated Mg^{2+} . We are investigating the structure of these electrolytes using NMR and single-crystal X-ray diffraction.

The cathode preparation involved high-temperature synthesis of CuMo_3S_4 from the elements¹⁶. Composite electrodes were then prepared from a mixture of this active mass with carbon and PVdF (polyvinylidene difluoride) binder (each 5–10 wt%; refs 6, 7), following electrochemical removal of the copper in an acetonitrile solution containing Mg salts. Alternatively, the active mass Mo_3S_4 could be prepared by chemical removal of the copper from CuMo_3S_4 using an aqueous FeCl_3 solution.

Mg can be inserted electrochemically into Mo_3S_4 up to a stoichiometry of MgMo_3S_4 (verified by element analysis based on atomic absorption and X-ray diffraction), corresponding to a charge capacity of $\sim 122 \text{ mA h g}^{-1}$, most of which can be reversibly utilized. The charge capacity of the new cathode is of the same order of magnitude as several important cathodes for Li batteries, such as a $\text{Li}_x\text{Mn}_2\text{O}_4$ spinel, which possesses a practical charge density of the order of $130\text{--}140 \text{ mA h g}^{-1}$ (ref. 6).

Standard coin cells (NRC) were used. The active mass comprised home-made, thin Mg foil anodes (10–100 μm thick), composite Mo_3S_4 cathodes containing additives $\sim 10 \text{ wt\%}$ (carbon and binder), and a glass cloth separator with a thin electrolyte solution layer.

Received 7 February; accepted 17 August 2000.

1. Lu, Z., Schechter, A., Moshkovich, M. & Aurbach, D. On the electroanalytical behavior of magnesium in a polar aprotic electrolyte solution. *J. Electroanal. Chem.* **466**, 203–217 (1999).
2. Gendler, J. D. & Pletcher, D. Studies using microelectrodes of the Mg^{++}/Mg couple in tetrahydrofuran and propylene carbonate. *J. Electroanal. Chem.* **199**, 93–100 (1986).
3. Brown, O. R. & McIntyre, R. The magnesium and magnesium amalgam electrodes in aprotic organic solvents. A kinetic study. *Electrochim. Acta* **30**, 627–633 (1985).
4. Novak, P., Tuhof, R. & Haas, O. Magnesium insertion electrodes of rechargeable nonaqueous batteries. A competitive alternative to lithium? *Electrochim. Acta* **45**, 351–367 (1999).
5. Aurbach, D., Moshkovich, M., Schechter, A. & Turgeman, R. The study of magnesium deposition and dissolution processes in ethereal Grignard salt solution using simultaneous EQCM-EIS and *in situ* FTIR spectroscopic measurements. *Electrochem. Solid State Lett.* **3**, 31–34 (2000).
6. Aurbach, D. *et al.* Common electroanalytical behavior of Li intercalation processes into graphite and transition metal oxides. *J. Electrochem. Soc.* **145**, 3024–3034 (1998).
7. Aurbach, D. *et al.* New insights into the interactions between electrode materials and electrolyte solutions for advanced nonaqueous batteries. *J. Power Sources* **81–82**, 95–111 (1999).
8. Connor, J. H., Reid, W. E. & Wood, G. B. Electrodeposition of metals from organic solutions: electrodeposition of magnesium and magnesium alloys. *J. Electrochem. Soc.* **104**, 38–41 (1957).
9. Liebenow, C. Reversibility of electrochemical magnesium deposition from Grignard salt solutions. *J. Appl. Chem.* **27**, 221–225 (1997).
10. Gregory, T., Hoffman, R. & Winterton, R. Nonaqueous electrochemistry of magnesium. Application to energy storage. *J. Electrochem. Soc.* **137**, 775–780 (1990).
11. Blomgren, J. in *Nonaqueous Electrochemistry* Ch. 2 (ed. Aurbach, D.) 53–79 (Dekker, New York, 1999).
12. Aurbach, D. & Moshkovich, M. The study of Li deposition–dissolution processes in a few selected electrolyte solutions by electrochemical quartz crystal microbalance (EQCM). *J. Electrochem. Soc.* **145**, 2629–2639 (1998).
13. Aurbach, D., Goren, E. & Chusid, O. The application of *in situ* FTIR spectroscopy to the study of surface films formed on lithium and noble metal at low potentials in Li battery electrolytes. *J. Electrochem. Soc.* **138**, L6–L9 (1991).
14. Aurbach, D., Weissman, I., Schechter, A. & Cohen, H. XPS studies of Li surfaces prepared in several important electrolyte solutions. A comparison with previous studies by FTIR spectroscopy. *Langmuir* **12**, 3991–4007 (1996).
15. Cohen, Y. & Aurbach, D. The use of a special work station for *in situ* measurements of highly reactive electrochemical systems by atomic force and scanning tunneling microscopes (AFM, STM). *Rev. Sci. Instrum.* **70**, 4668–4675 (1999).

16. Yvon, K. in *Current Topics in Material Science* Vol. 3 (ed. Kaldis, E.) 53–129 (North-Holland, Amsterdam, 1979).
17. Chevrel, R., Sergent, M. & Prigent, J. Sur le nouvelles phases sulfurées ternaires du Molybdène. *J. Solid State Chem.* **3**, 515–519 (1971).
18. Ritter, C., Gocke, E., Fischer, C. & Schollhorn, R. Neutron diffraction study of the crystal structure of lithium intercalated Chevrel phases. *Mater. Res. Bull.* **27**, 1217–1225 (1992).
19. Linden, D. in *Handbook of Batteries* 2nd edn Ch. 23 (ed. Linden, D.) 23.3–23.22 (McGraw Hill, New York, 1994).

Acknowledgements

This work was partially supported by Advance Technology Upgrading (ATU) Ltd, Israel.

Correspondence and requests for materials should be addressed to D.A.
(e-mail: aurbach@mail.biu.ac.il).

Importance of stirring in the development of an iron-fertilized phytoplankton bloom

Edward R. Abraham*, Cliff S. Law†, Philip W. Boyd‡, Samantha J. Lavender†§, Maria T. Maldonado||§ & Andrew R. Bowie#†

* National Institute for Water and Atmospheric Research (NIWA),
PO Box 14-901, Kilbirnie, Wellington, New Zealand

† Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth,
Devon PL1 3DH, UK

‡ NIWA Centre for Chemical and Physical Oceanography, Department of
Chemistry, University of Otago, Dunedin, New Zealand

|| Department of Biology, McGill University, 1205 Dr. Penfield Avenue, Montreal,
PQ H2T 2V8, Canada

Department of Environmental Sciences, Plymouth Environmental Research
Centre, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

The growth of populations is known to be influenced by dispersal, which has often been described as purely diffusive^{1,2}. In the open ocean, however, the tendrils and filaments of phytoplankton populations provide evidence for dispersal by stirring^{3,4}. Despite the apparent importance of horizontal stirring for plankton ecology, this process remains poorly characterized. Here we investigate the development of a discrete phytoplankton bloom, which was initiated by the iron fertilization of a patch of water (7 km in diameter) in the Southern Ocean⁵. Satellite images show a striking, 150-km-long bloom near the experimental site, six weeks after the initial fertilization. We argue that the ribbon-like bloom was produced from the fertilized patch through stirring, growth and diffusion, and we derive an estimate of the stirring rate. In this case, stirring acts as an important control on bloom development, mixing phytoplankton and iron out of the patch, but also entraining silicate. This may have prevented the onset of silicate limitation, and so allowed the bloom to continue for as long as there was sufficient iron. Stirring in the ocean is likely to be variable, so blooms that are initially similar may develop very differently.

Remotely sensed ocean colour images show that there was an intense bloom in the vicinity of the SOIREE (Southern Ocean iron release experiment) site during March 1999 (Fig. 1). The phytoplankton bloom seen on 23 March has a peak chlorophyll *a* (chl*a*) concentration of 3 mg m⁻³ (see Methods), which is extraordinary for this region^{6,7}, particularly as March is towards the end of the

growth season. To confirm that the feature was unusual, we analysed all the daily composite SeaWiFS data collected during the 1997–2000 September to March period, over the area between 134–146° E and 57–62° S. The mean concentration was 0.20 ± 0.06 mg chl*a* m⁻³, and, apart from the images shown in Fig. 1 and occasional isolated pixels, there were no chlorophyll *a* concentrations greater than 1 mg m⁻³. Here we argue that the remotely sensed bloom is derived from the fertilization experiment. We show that stirring and diffusion of the SOIREE patch could have produced a filament of the dimensions seen in the images, and that there was sufficient iron in the fertilized patch to support the bloom during March.

The stretching of a tracer filament allows the strength of the stirring in a two-dimensional flow to be estimated (see Methods). A transect was made on 18 February, which showed that the patch which was responding to the fertilization was 30 km long (as evidenced by enhanced photosynthetic competency, $F_v/F_m > 0.3$). The only remotely sensed image of the entire bloom is from 23 March (Fig. 1d); by this time the patch appears as a 150-km-long ribbon of high chlorophyll *a* concentration. The change in the length of the patch over the intervening month (33 d) implies an effective strain rate of $\gamma = \ln(150/30)/33 = 0.05$ d⁻¹ = 6×10^{-7} s⁻¹. This is a minimum value, as the phytoplankton at the margins of the patch may be iron-limited, which would lead to the true length of the patch on 23 March being underestimated. By fitting gaussian profiles to the chlorophyll *a* concentration along cross-filament sections, the width of the bloom in the 23 March image is found to be $2\sigma_y = 4 \pm 2$ km. This gives a diffusivity, $K_y = \sigma_y^2 \gamma$, of between 0.5 and 5 m² s⁻¹. These values of γ and K_y may be compared with estimates obtained from changes in the size and shape of the patch over the fortnight that it was followed by ship (Fig. 2). The patch's length increased over the course of the experiment, whereas its width stabilized at 4 km, the same as the width of the filament in the satellite images. The best fit to the length and width data is given by $\gamma = (8 \pm 3) \times 10^{-7}$ s⁻¹ and $K_y = 4 \pm 2$ m² s⁻¹, which agree very well with the values above. Here we use $\gamma = 8 \times 10^{-7}$ s⁻¹ = 0.07 day⁻¹ as the best estimate of the rate of stretching of the bloom.

The measured effective strain rate is close to the r.m.s. strain of 5.8×10^{-7} s⁻¹ derived from a generic model of mesoscale turbulence⁸, and is of similar magnitude to the effective strain rate of $(3 \pm 0.5) \times 10^{-7}$ s⁻¹ determined by a tracer release within the ocean interior⁹. The horizontal diffusivity is lower than values found in previous open-ocean surface SF₆ releases: namely, 22 ± 10 m² s⁻¹ within an eddy in the North Atlantic Ocean¹⁰ and 25 ± 5 m² s⁻¹ in the equatorial Pacific Ocean (recalculated from ref. 11). Our value of the horizontal diffusivity is, however, within the range (2–16 m² s⁻¹) found at the length-scale $3\sqrt{2}\sigma_y = 8$ km from dye-release experiments in shelf seas¹². The plausibility of our derived values supports the assertion that the remotely sensed bloom was produced by the fertilization experiment.

The presence of horizontal stirring has implications for the growth of phytoplankton in an isolated patch. Stretching by the horizontal currents maintains cross-patch gradients against diffusion, causing continuing mixing along the sides of the filament. This results in a persistent dilution of the concentration at the patch centre¹³. If the net phytoplankton growth rate, μ , is higher than the strain rate, then the peak phytoplankton concentration within the patch will increase at the asymptotic rate $\mu - \gamma$ (see Methods). The carbon-based phytoplankton growth rate reached 0.19 d⁻¹ by the time the ship left the SOIREE site⁵ and the strain rate was $\gamma = 0.07$ d⁻¹, so towards the end of the site occupation approximately 35% of the daily production was lost from the patch centre through horizontal dispersal. The losses due to dispersal were much higher than those from meso-zooplankton grazing (less than 0.01 d⁻¹, J. R. Zeldis, personal communication) and vertical sinking (less than 0.02 d⁻¹ (ref. 14)). Stirring rates are expected to be variable¹⁵, and are likely to be higher in regions with more eddy

§ Present addresses: Institute of Marine Studies, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK (S.J.L.); School of Marine Sciences, 5741 Libby Hall, University of Maine, Orono, Maine 04469, USA (M.M.)

activity than the SOIREE site, which was chosen for its weak surface currents⁵. If the strain rate is higher than the growth rate, then, from equation (2) in Methods, the phytoplankton concentration in an isolated bloom will never grow beyond $\gamma C_0/(\gamma - \mu)$, where C_0 is the background concentration. The importance of stirring poses a potential difficulty for iron fertilization. If a fertilized patch happened to be stirred rapidly then there would be little increase in phytoplankton stocks, despite the phytoplankton no longer being iron-limited. Because particle export from the surface layer has a strongly nonlinear dependence on algal concentration¹⁶, this may result in the added iron failing to induce a significant phytoplankton-mediated removal of carbon from the surface waters.

The iron fertilization caused the accumulation of between 600 and 3,000 t of algal carbon by the time of the 23 March image (see

Methods). On-deck incubations measured an intracellular phytoplankton uptake ratio of 3×10^{-6} mol Fe per mol C (the mean from days 5 and 12 of SOIREE was $(2.7 \pm 1.4) \times 10^{-6}$ mol Fe per mol C, for the $>2 \mu\text{m}$ phytoplankton size fraction, see Methods). If this uptake ratio remained constant, then the phytoplankton in the 23 March bloom would have contained between 8 and 40 kg of iron, an order of magnitude less than the 300 kg of dissolved iron which remained in the patch after the last infusion. There was a dissolved iron concentration of 1.1 nM at the centre of the fertilized patch at the end of the site occupation, and a background concentration of 0.1 nM. By this time there was an excess of organic iron-binding ligands⁵, maintaining the iron in solution. Towards the end of the site occupation there were minimal losses of dissolved iron from the surface layer, with the decrease in concentration being similar to what would be expected from horizontal

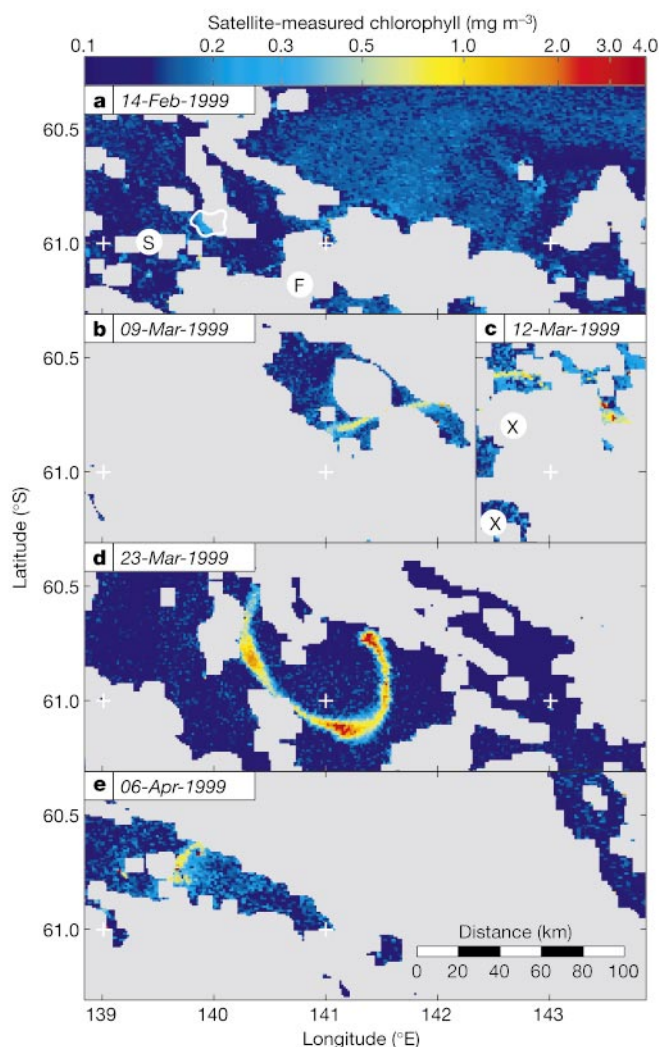


Figure 1 Images of the SOIREE bloom from the SeaWiFS ocean colour satellite. The region is usually cloudy, so during the months February–April only eight images of the bloom were obtained. Five of these are displayed here. The images show sea-surface chlorophyll *a* concentration on a logarithmic scale (see Methods), and areas that were obscured by cloud are marked in grey. Composite images of the bloom may be viewed at <http://seawifs.gsfc.nasa.gov>; the bloom is clearly visible in the March monthly climatology (1998–2000). **a**, The only image taken during the site occupation. S indicates the start of the SOIREE experiment, on 9 February 1999, and F indicates the position of the patch centre at the time of the final survey on 22 February. The white line is the edge of the patch, based on a shipboard survey at the time the image was taken. The chlorophyll *a* concentration within the contour is 0.1 mg m^{-3} higher than the background level.

Phytoplankton concentrations within the patch reached $2 \text{ mg chl a m}^{-3}$ by the time of the final survey. **b, c**, By 1 month after the initial release, the patch has become drawn out into a long streak. Another partial image was obtained on 7 March, which shows that the bloom extends no further to the west than seen here. X symbols mark the positions of temperature profiles taken on 3 March, which showed that the mixed layer was then 65 m deep. **d**, After 42 d, the fertilized patch has become stretched and folded into a 150-km-long ribbon. The area of the bloom is $1,100 \text{ km}^2$ ($>0.2 \text{ mg chl a m}^{-3}$), with the high-chlorophyll portion having an area of 230 km^2 ($>1.0 \text{ mg chl a m}^{-3}$). Two other partial images from 21 and 22 March show that this is close to being a complete image of the bloom. **e**, The bloom is still present 55 d after the initial release. This is the last image of the bloom that was obtained.

dispersal alone. The concentration of dissolved iron remaining in the patch centre at the time of the satellite images may then be calculated under the assumption that, apart from uptake required to maintain net phytoplankton growth, the dissolved iron was conserved. To keep the phytoplankton concentration steady at $2 \text{ mg chl } a \text{ m}^{-3}$, a growth rate of 0.10 d^{-1} (0.07 d^{-1} for dispersal and 0.03 d^{-1} to cover other losses such as sinking, which may not result in the iron being recycled), or $0.2 \text{ mg chl } a \text{ m}^{-3} \text{ d}^{-1}$, would have been required. Given an intermediate carbon-to-chlorophyll *a* ratio of $60 \text{ g C per g chl } a$, and the iron-to-carbon uptake ratio quoted above, only 3 pM d^{-1} of iron would have been needed by the phytoplankton at the bloom centre. It follows that there would still have been a peak concentration of 0.2 nM of dissolved iron within the filament on 23 March (see Methods), double the background concentration and sufficient to prevent severe iron limitation¹⁷. Although this calculation relies on the assumption that dissolved iron was retained within the surface layer, as was observed towards the end of the site occupation, a far higher loss rate of 11 pM d^{-1} would have been needed to reduce the concentration to background levels before 23 March. This suggests that, despite the stirring, there was enough dissolved iron in the fertilized patch to maintain increased phytoplankton growth during March.

The predominant taxa in the bloom when the ship left the SOIREE site were diatoms, which have a high silicate requirement, and the silicate concentration in the mixed layer decreased from 10 to $6 \text{ } \mu\text{M}$ during the 12-day site occupation. Although succession by other taxa is perhaps possible, phytoplankton blooms in open Southern Ocean waters are typically dominated by diatoms¹⁸, and so the development of the bloom for a further month suggests that stirring of the patch entrained enough silicate to support continued diatom growth. An algal uptake ratio of $0.18 \text{ mol Si per mol C}$ was measured during SOIREE, so a silicate uptake of $0.18 \text{ } \mu\text{M d}^{-1}$ would have been needed to maintain a growth rate of $0.2 \text{ mg chl } a \text{ m}^{-3} \text{ d}^{-1}$. The silicate drawdown can be estimated as the uptake rate divided by the strain rate; so a drawdown of $0.18/0.07 = 2.6 \text{ } \mu\text{M}$ is all that was required to supply sufficient silicate to the bloom. While the filament was being stirred there would not have been any silicate limitation, which helps to explain its longevity. An input of iron that was spread over an area with dimensions larger than the $\sim 100\text{-km}$ eddy length scale may result in a relatively short-lived bloom, as above this scale stirring is inefficient. Depending on their initial

concentration, other nutrients such as silicate may then become limiting, leading to the decline of the bloom¹⁹.

We have presented the first (to our knowledge) experimental demonstration of the mesoscale stirring of a tracer patch in the surface ocean, a result made possible by the visibility of phytoplankton in satellite images. The stretching of the patch into a long filament agrees with recent theoretical analysis¹³, and supports the idea that the tendrils and whorls seen in naturally occurring plankton populations may be generated through horizontal stirring within the surface layer^{3,20,21}. The calculations presented here could be refined through the use of a numerical model that included a detailed representation of phytoplankton and iron dynamics. It is clear, however, that algal losses due to dispersal may dominate those from other processes, and that horizontal stirring will affect the concentration of nutrients within an isolated patch or filament. For these reasons, a consideration of stirring is essential to understanding the development of phytoplankton blooms in the open ocean. □

Methods

Satellite imagery

The SeaWiFS High-Resolution Picture Transmission (1-km resolution) data was obtained from the NASA GDAAC and processed using SeaAPS²². The images presented are level 2 chlorophyll *a* concentrations, estimated using the Ocean Chlorophyll 2 version 2 (OC2-v2) algorithm $\text{chl } a \text{ (mg m}^{-3}\text{)} = 10^{(0.2974 - 2.2429x + 0.8358x^2 - 0.0077x^3) - 0.0929}$, where x is the $490/555 \text{ nm}$ band ratio²³. The algorithm has not been extensively calibrated against *in situ* extracted-chlorophyll data from the Southern Ocean, although recent results suggest that it underestimates chlorophyll *a* concentrations there²⁴.

Dispersal of a tracer in the surface ocean

At scales greater than 1 km , and away from strong fronts, the currents in the surface ocean are approximately two-dimensional and divergence-free. The flow can then be locally resolved into a pure strain and a rotation. In a pure strain flow with a velocity $\mathbf{v} = (\gamma x, -\gamma y)$, where γ is the strain rate, a patch of tracer is stretched along the x -axis and squeezed in the y -direction. After an initial period of adjustment, an isolated patch of conserved tracer, $C(x, y, t)$, which is centred at the origin, becomes a gaussian ellipsoid;

$$C = (N/2\pi\sigma_x\sigma_y)\exp(-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})) \quad (1)$$

where N is the total amount of tracer within the patch and $\sigma_x(t)$, $\sigma_y(t)$ are the standard deviations of the distribution in the x and y directions, respectively. The length of the patch increases exponentially at the rate γ , $\sigma_x(t) \propto \exp(\gamma t)$. In contrast, the width stabilizes at a steady value where the thinning effect of the strain balances the widening tendency of diffusion, $\sigma_y = (K/\gamma)^{1/2}$ (refs 9, 25). Once the width has stabilized, the peak concentration decreases at the rate γ . In ocean flows the strain is not constant, but an effective strain rate (more formally, a Lyapunov exponent) may be defined from the exponential growth in the length of a patch of tracer^{15,26}. Once the length of the filament becomes larger than the sizes of the eddies, the patch tends to be folded back upon itself and the tracer distribution loses its streakiness²⁶.

A pure-strain flow stretches a phytoplankton patch into a filament of the same dimensions as a patch of conserved tracer¹³. The change in the peak phytoplankton concentration in a fertilized bloom may be approximated by $dC/dt = \mu C - \gamma(C - C_0)$, where μ is the net growth rate the phytoplankton within the fertilized patch would have in the absence of dispersal; C_0 is the background concentration, which is assumed to be constant, and the second term represents losses due to dispersal. This equation has the solution

$$C(t) = \frac{\mu}{\mu - \gamma} C_0 \exp((\mu - \gamma)t) - \frac{\gamma}{\mu - \gamma} C_0 \quad (\mu \neq \gamma), \quad C(t) = C_0 + \mu C_0 t \quad (\mu = \gamma) \quad (2)$$

Algal carbon within the bloom

By integrating over the portion of the 23 March image (Fig. 1d) with concentrations greater than $0.2 \text{ mg chl } a \text{ m}^{-3}$, the near-surface chlorophyll *a* content of the bloom, per unit depth, is found to be 820 kg m^{-1} . There was an expendable bathythermograph (XBT) survey through the region of the bloom in early March (Fig. 1c), which showed that the surface waters were well mixed to a depth of 65 m , consistent with that found during the site occupation. It is unlikely that the mixed layer shoaled during March²⁷, and so the remotely sensed bloom can be assumed to have a 65-m extent. Given that the carbon-to-chlorophyll ratios remained in the range measured during SOIREE ($45\text{--}90 \text{ g C per g chl } a$), the 23 March bloom is estimated to have contained between $2,400$ and $4,800 \text{ t}$ of algal carbon. Areal algal carbon outside the fertilized patch was 1.6 g m^{-2} during the site occupation. If this remained steady, then there would have been $1,800 \text{ t}$ of algal carbon in the bloom waters in the absence of iron fertilization, implying that the fertilization caused the accumulation of between 600 and $3,000 \text{ t}$ of algal carbon by 23 March. Because the calibration of the SeaWiFS data is uncertain, the carbon-to-chlorophyll ratios are

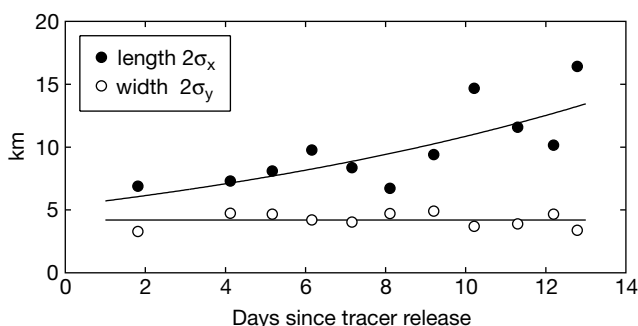


Figure 2 Changes in the size and shape of the fertilized patch following the initial infusion. A tracer, SF_6 , was added to the patch during the initial iron release⁵, enabling the patch to be mapped using an underway system³⁰. The width and length of the patch were determined by fitting a gaussian ellipsoid (equation (1) in Methods) to the SF_6 data from each daily survey, using a least-squares fitting procedure. The r.m.s. difference between the data and the best-fit ellipsoid was less than 15% of the peak SF_6 concentration for all the points shown. A best fit to the length and width of the patch after day 4 ($2\sigma_x \propto \exp(\gamma t)$, $2\sigma_y = 2(K/\gamma)^{1/2}$; shown by the lines) was used to estimate the strain rate and the diffusivity. The size of the patch that had above-background SF_6 concentrations was larger than the size obtained using the method described here. The patch defined by SF_6 concentrations above 8 fM reached a length of 30 km by 18 February (9 d after the initial release).

estimated, and the true vertical structure of the bloom is unknown, these figures must be treated with caution, and they may be subject to revision.

Iron concentration within the patch

The rate of change of the dissolved iron concentration at the patch centre may be estimated as $d\text{Fe}_{\text{max}}/dt = -\gamma(\text{Fe}_{\text{max}} - \text{Fe}_{\text{out}}) - P$, where Fe_{max} is the peak concentration within the patch; Fe_{out} is the background concentration; and P is a constant loss term, assumed to be the net rate of phytoplankton iron uptake. This equation has the solution $\text{Fe}_{\text{max}}(t) = (\text{Fe}_{\text{initial}} - \text{Fe}_{\text{out}} + P/\gamma) \exp(-\gamma t) + \text{Fe}_{\text{out}} - P/\gamma$, where $\text{Fe}_{\text{initial}}$ is the peak iron concentration at $t = 0$.

The phytoplankton iron-uptake rate was obtained from the estimated net rate of carbon fixation and the Fe:C uptake ratio, which was measured by parallel incubation of ^{55}Fe ($0.2 \text{ nM } ^{55}\text{Fe}$: $10 \text{ } \mu\text{M EDTA}$) and $\text{H}^{14}\text{CO}_3^-$, as described previously²⁸. The Fe and EDTA addition resulted in a concentration of inorganic Fe (Fe') of 7.7 pM which mimicked the mean *in situ* Fe' concentration ($4.5 \pm 2.4 \text{ pM}$; P. L. Croft, personal communication). Underway dissolved iron was measured using a flow injection–chemiluminescence method²⁹.

With the values (see text) $\text{Fe}_{\text{initial}} = 1.1 \text{ nM}$, $\text{Fe}_{\text{out}} = 0.1 \text{ nM}$, $P = 0.003 \text{ nM d}^{-1}$, and $\gamma = 0.07 \text{ d}^{-1}$ it follows that after 28 d, $\text{Fe}_{\text{max}} = 0.20 \text{ nM}$. If the loss rate is $P = 0.011 \text{ nM d}^{-1}$ then after 28 d, $\text{Fe}_{\text{max}} = 0.10 \text{ nM}$.

Received 20 December 1999; accepted 4 September 2000.

- Kierstead, H. & Slobodkin, L. B. The size of water masses containing plankton blooms. *J. Mar. Res.* **12**, 141–147 (1953).
- Okubo, A. *Biomathematics Vol. 10, Diffusion and Ecological Problems* (Springer, Berlin, 1980).
- Gower, J. F. R., Denman, K. L. & Holyer, R. J. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. *Nature* **288**, 157–159 (1980).
- Holligan, P. *et al.* A biogeochemical study of the coccolithophore, *Emiliania Huxleyi*, in the North Atlantic. *Glob. Biogeochem. Cycles* **7**, 879–900 (1993).
- Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–701 (2000).
- Comiso, J. C., McClain, C. R., Sullivan, C. W., Ryan, J. P. & Leonard, C. L. Coastal Zone Color Scanner pigment concentrations in the southern ocean and relationships to geophysical surface features. *J. Geophys. Res.* **C 2**, 2419–2451 (1993).
- Sullivan, C. W. *et al.* Distributions of phytoplankton blooms in the southern ocean. *Science* **262**, 1832–1837 (1993).
- Haidvogel, D. B. & Keffer, T. Tracer dispersal by mid-ocean mesoscale eddies. Part I. Ensemble statistics. *Dyn. Atmos. Oceans* **8**, 1–40 (1984).
- Ledwell, J. R., Watson, A. J. & Law, C. S. Mixing of a tracer in the pycnocline. *J. Geophys. Res.* **C 103**, 21499–21529 (1998).
- Martin, A. P., Richards, K. J., Law, C.-S. & Liddicoat, M. I. Horizontal dispersion within an anticyclonic mesoscale eddy. *Deep-Sea Res.* (in the press).
- Stanton, T. P., Law, C. S. & Watson, A. J. Physical evolution of the IronEx-I open ocean tracer patch. *Deep-sea Res.* **II 45**, 947–975 (1998).
- Okubo, A. Oceanic diffusion diagrams. *Deep-Sea Res.* **18**, 789–802 (1971).
- Martin, A. P. On filament width in oceanic plankton populations. *J. Plank. Res.* **22**, 597–602 (2000).
- Charette, M. A. & Buessler, K. O. Does iron fertilization lead to rapid carbon export in the Southern Ocean? *Geochem. Geophys. Geosyst.* [online] **1**, Paper number 2000GC000609 (<http://www.g-cubed.org>) (2000).
- Sundermeyer, M. A. & Price, J. F. Lateral mixing and the North Atlantic Tracer Release Experiment: observations and simulations of lagrangian particles and a passive tracer. *J. Geophys. Res.* **103**, 21481–21497 (1998).
- Jackson, G. A. A model of the formation of marine algal flocs by physical coagulation processes. *Deep-Sea Res.* **37**, 1197–1211 (1990).
- Olson, R. J., Sosik, H. M., Chekalyuk, A. M. & Shalapyonok, A. Effects of iron enrichment on phytoplankton in the Southern Ocean during late summer: Active fluorescence and flow cytometric analyses. *Deep-Sea Res.* (in the press).
- de Baar, H. J. W. & Boyd, P. W. in *The Dynamic Ocean Carbon Cycle: A Midterm Synthesis of the Joint Global Ocean Flux Study* (eds Hanson, R. B., Ducklow, H. W. & Field, J. G.) 61–140 (International Geosphere Biosphere Programme Book Series, Cambridge Univ. Press, Cambridge, 1999).
- Abbott, M. R., Richman, J. G., Letelier, R. M. & Bartlett, J. S. The spring bloom in the Antarctic Polar Frontal Zone as observed from a mesoscale array of bio-optical sensors. *Deep-Sea Res.* (in the press).
- Washburn, L., Emery, B. M., Jones, B. H. & Ondercin, D. G. Eddy stirring and phytoplankton patchiness in the subarctic North Atlantic in late summer. *Deep-Sea Res.* **I 45**, 1411–1439 (1998).
- Abraham, E. R. The generation of plankton patchiness by turbulent stirring. *Nature* **391**, 577–580 (1998).
- Lavender, S. J. & Groom, S. B. The SeaWiFS Automatic Data Processing System (SeaAPS). *Int. J. Remote Sensing* **20**, 1051–1056 (1999).
- O'Reilly, J. E. *et al.* Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.* **C 103**, 24937–24953 (1998).
- Moore, J. K. *et al.* SeaWiFS satellite ocean color data from the Southern Ocean. *Geophys. Res. Lett.* **26**, 1465–1468 (1999).
- Okubo, A. A note on horizontal diffusion from an instantaneous source in a nonuniform flow. *J. Oceanogr. Soc. J.* **22**, 1–6 (1966).
- Garrett, C. On the initial streakiness of a dispersing tracer in two- and three-dimensional turbulence. *Dyn. Atmos. Oceans* **7**, 265–277 (1983).
- Rintoul, S. R., Donguy, J. R. & Roemmich, D. H. Seasonal evolution of upper ocean thermal structure between Tasmania and Antarctica. *Deep-Sea Res.* **I 44**, 1185–1202 (1997).
- Maldonado, M. T., Boyd, P. W., Harrison, P. J. & Price, N. M. Co-limitation of phytoplankton growth by light and Fe during winter in the NE subarctic Pacific Ocean. *Deep-Sea Res.* **II 46**, 2475–2485 (1999).

- Bowie, A. R., Achterberg, E. P., Mantoura, R. F. C. & Worsfold, P. J. Determination of sub-nanomolar levels of iron in seawater using flow injection with chemiluminescence detection. *Anal. Chim. Acta* **361**, 189–200 (1998).
- Law, C. S., Watson, A. J., Liddicoat, M. I. & Stanton, T. Sulphur hexafluoride as a tracer of biogeochemical and physical processes in an open-ocean iron fertilisation experiment. *Deep-Sea Res.* **II 45**, 977–994 (1998).

Acknowledgements

We thank all the SOIRE team, particularly the captain and crew of RV *Tangaroa* for their assistance throughout the experiment; S. Nodder, R. Murdoch, A. Watson and T. Trull for experiment planning and coordination; and G. Jameson, M. Liddicoat and R. Ling for their work on the tracer release and mapping. We also thank S. Groom for additional interpretation of the remote sensing data; P. Nightingale for ARGOS data transfer; S. Rintoul for providing the XBT data; and M. Morris, C. Stevens and A. Martin for comments on the original manuscript. The SeaWiFS data, provided by the NASA DAAC/GSF and copyright of Orbital Imaging Corps and the NASA SeaWiFS project, was processed at CCMS-PML. E.A. and P.B. acknowledge the financial assistance of the NZ Public Good Science Fund for Antarctic research; C.L. thanks the UK Natural Environment Research Council for support; M.M. was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Center for Environmental Bioinorganic Chemistry, Princeton (CEBIC); A.B. was supported by the University of Plymouth and Plymouth Marine Laboratory.

Correspondence and requests for materials should be addressed to E.A. (e-mail: e.abraham@niwa.cri.nz).

Effect of iron supply on Southern Ocean CO₂ uptake and implications for glacial atmospheric CO₂

A. J. Watson*, D. C. E. Bakker*, A. J. Ridgwell*, P. W. Boyd† & C. S. Law‡

* School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

† NIWA Centre for Chemical and Physical Oceanography, Department of Chemistry, University of Otago, Dunedin, New Zealand

‡ Centre for Coastal and Marine Studies, Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

Photosynthesis by marine phytoplankton in the Southern Ocean, and the associated uptake of carbon, is thought to be currently limited by the availability of iron^{1,2}. One implication of this limitation is that a larger iron supply to the region in glacial times³ could have stimulated algal photosynthesis, leading to lower concentrations of atmospheric CO₂. Similarly, it has been proposed that artificial iron fertilization of the oceans might increase future carbon sequestration. Here we report data from a whole-ecosystem test of the iron-limitation hypothesis in the Southern Ocean⁴, which show that surface uptake of atmospheric CO₂ and uptake ratios of silica to carbon by phytoplankton were strongly influenced by nanomolar increases of iron concentration. We use these results to inform a model of global carbon and ocean nutrients, forced with atmospheric iron fluxes to the region derived from the Vostok³ ice-core dust record. During glacial periods, predicted magnitudes and timings of atmospheric CO₂ changes match ice-core records well. At glacial terminations, the model suggests that forcing of Southern Ocean biota by iron caused the initial ~40 p.p.m. of glacial–interglacial CO₂ change, but other mechanisms must have accounted for the remaining 40 p.p.m. increase. The experiment also confirms that modest sequestration of atmospheric CO₂ by artificial additions of iron to the Southern Ocean is in principle possible, although the period and geographical extent over which sequestration would be effective remain poorly known.

The *in situ* iron-enrichment experiment (SOIREE) was performed at 61°S, 141°W in February 1999⁴. The method was that of a tracer-steered iron release first described in ref. 5 and previously employed in the equatorial Pacific Ocean^{6,7}. A solution of ferrous sulphate in acidified sea water was released over a period of about 12 h, in a region centred on a drogued drifting buoy, to create a patch of surface water—of linear dimensions approximately 10 km—enriched to a dissolved iron concentration of about 3 nM. Sulphur hexafluoride (SF₆) tracer was released in a known ratio to iron in the initial release, enabling the subsequent advection and dilution of the patch to be accounted for. On days 4, 6 and 8 after the start of the experiment, when dissolved iron concentrations in the patch fell below about 0.3 nM, further additions of iron to the centre of the tracer-marked region were made.

Figure 1 shows the time sequence of surface water measurements of CO₂ fugacity (f_{CO_2}) recorded on board the survey vessel during the 13 days of the experiment, as it moved continuously in and out of the enriched patch. The instrument and data reduction techniques used were similar to those described in ref. 8. The increase with time, and much of the variability, of the “out of patch” f_{CO_2} is largely accounted for by changes in temperature of the mixed layer, which warmed about 0.3 °C during the period. The iron release changed the water from being approximately neutral with respect to atmospheric CO₂ to a sink, with surface water f_{CO_2} being 25–30 μatm below the atmospheric value in the patch at the end of the period. These are the first, to our knowledge, whole-ecosystem data that show that biologically mediated drawdown of surface f_{CO_2} responds strongly to the availability of iron in the polar Southern Ocean. Satellite observation of a feature—which we identify as the aftermath of the experiment six weeks after initial release—shows chlorophyll at concentrations similar to those that prevailed when we left the site at the end of two weeks, but in a patch covering 5–10 times the area at that time⁹. This suggests that the sink for atmospheric CO₂ created by the iron release continued to grow in magnitude after we left the area.

The ratio of silicon to carbon fixed by the ecosystem is important for interpretations of the effect of iron fluxes on atmospheric CO₂ concentrations in glacial times, as indicated by model results reported here (see below, and Supplementary Information). Our

measurements of the ratio of net uptake of silicon to carbon by the ecosystem showed a substantial response to iron addition, with the amount of silicon fixed per mole of carbon inside the patch being only about half that outside (see Supplementary Information). The ratio of uptake in the patch was established both by isotopic uptake incubations, and by examination of changes in bulk silicon and carbon concentration in the patch. Incubations taken from 5 m depth show mean molar uptake ratios Si:C = 0.09 after day 5 of the experiment, while outside the patch the ratio was 0.17. Our observations are consistent with those recently reported in bottle iron-enrichment experiments in the Antarctic¹⁰ and elsewhere¹¹. When account is taken of the different variations with depth in the mixed layer of the silicon and carbon uptake rates (see Supplementary Information), we estimate 0.19 and 0.35 for respectively inside- and outside-patch Si:C over the mixed layer as a whole (these values are similar to the *in vitro* observations). We believe that ours are the first measurements to show an effect of iron on Si:C uptake ratios in the unenclosed marine ecosystem.

What is the source of the iron needed by the Southern Ocean biota? It has generally been assumed^{1,12} that local atmospheric dust is the main source, but recent work has suggested instead that dissolved iron in upwelling water is more important¹³. Recent measurements indicate that the concentrations of dissolved iron in the Southern Ocean at depths of 800 m are about 0.3 nM far from land¹⁴, while surface water concentrations are limiting at about 0.05–0.1 nM. ¹⁴C tracer budgets¹⁵ and circulation models¹⁶ are in agreement about the magnitude of the upwelling flux of water (approximately 20–40 Sv). From these figures, we calculate a supply of iron by upwelling averaging 8–16 $\mu\text{mol m}^{-2} \text{yr}^{-1}$ for the area (approximately $2 \times 10^{13} \text{ m}^2$) south of the polar front. Estimates of the amount of iron supplied by atmospheric dust vary. Both the dust flux itself, and the fraction of dissolvable iron it contains, are poorly constrained. Following data and assumptions given in ref. 12, a flux of 0.2 $\mu\text{mol m}^{-2} \text{yr}^{-1}$ dissolvable Fe for the Antarctic region has been calculated¹³: 10% of the iron in the dust was assumed to be soluble. A recent modelling estimate of atmospheric dust fluxes¹⁷ suggests a dust flux to the region about 25 times higher than this. However, recent work¹⁸ suggests that the estimate of 10% of the iron in the dust being soluble is an order-of-magnitude overestimate. Thus, present-day upwelling flux appears to be on average larger than the atmospheric flux, probably by an order of magnitude.

The extent of surface depletion of macro-nutrients is consistent with upwelled iron being the limiting nutrient in Southern Ocean waters. Laboratory measurement of the Fe requirements of open-ocean phytoplankton¹⁹ show that, under conditions in which oceanic diatoms are severely iron-stressed, their molar C:Fe uptake ratios rise to a maximum of about $3 \times 10^5:1$. This ratio implies that the maximum net drawdown of inorganic carbon due to iron-stressed plankton removing about 0.25 nM Fe from upwelled water should be approximately 75 $\mu\text{mol kg}^{-1}$ C, or (using C:N ≈ 7) about 10 $\mu\text{mol kg}^{-1}$ nitrate. Nitrate concentrations in the region are indeed typically 10 $\mu\text{mol kg}^{-1}$ lower at the surface than at 800 m depth²⁰.

Ice cores³, sediments² and models¹⁷ indicate that iron fluxes from the atmosphere were higher in glacial times by a factor of 2–5 globally, and by a factor of 10–50 in the Southern Ocean. Therefore, during the glaciations, Southern Ocean atmospheric fluxes might have exceeded the upwelling source. This would result in a greater efficiency of the biological pump in the region. A similar increase in efficiency would be expected if the concentration of upwelling iron increased in response to global increases in the atmospheric source¹³.

Thus we might expect increased depletion of macro-nutrients during glacial times in this region, which could contribute to reduced atmospheric CO₂ concentrations²¹. To investigate quantitatively the effect of Southern Ocean dust flux on these variables, we have adapted a simple model of the ocean–atmosphere carbon

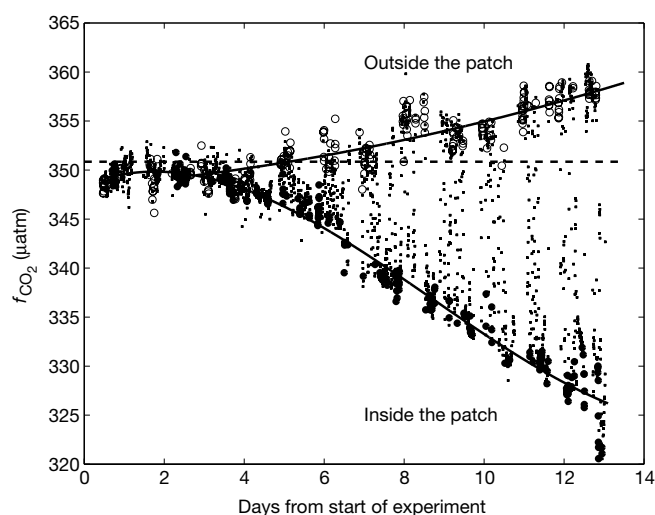


Figure 1 Seawater f_{CO_2} in surface water measured from the RV *Tangaroa* as a function of time during the experiment. The simultaneously measured SF₆ tracer concentrations were used to construct the heavy lines, as follows. “Outside the patch” shows a quadratic fit to the set of data points having the lowest 10% of SF₆ concentrations (open circles), while “Inside the patch” is a cubic fitted to the set of points having the highest 10% SF₆ concentrations on each day (filled circles). The dashed line shows the mean of atmospheric f_{CO_2} measured during this period.

cycle, incorporating measured effects of iron on f_{CO_2} and silica-carbon ratios (see Methods and Supplementary Information). Figure 2 shows the result of driving the model with a variable dust deposition rate to the Southern Ocean, with all other model boundary conditions held constant; the variable dust deposition rate was based on the dust concentration record in the Vostok core. Variations in model atmospheric CO_2 during glaciated periods and in the first part of the four “terminations” match the ice-core record well, as does the timing and approximate magnitude of the decrease in atmospheric $\delta^{13}\text{C}$ in the initial stage of termination I²² (not shown). Contrary to recent arguments²³ that the delay between the decline in dust flux and the change in atmospheric CO_2 concentrations at termination II was too long for the two to be directly linked by the Southern Ocean, we find the timings of events to be well matched. As a result of a nonlinear response of the system to changes in Fe supply, modelled CO_2 does not rise significantly until much of the decline in dust flux has taken place. The behaviour is a result of a shift from a glacial regime that is predominantly silicon-limited to a modern one that is predominantly iron-limited. The timing is sensitive to the forcing by dust flux, so that if the assumed present-day dust supply and/or the maximum glacial-interglacial change are significantly reduced, correlation between modelled and observed CO_2 signals at the terminations is degraded, whereas if we had assumed a higher solubility for iron in dust we would have obtained a better match.

The maximum model $[\text{PO}_4^{3-}]$ drawdown in the Southern Ocean is about $0.6 \mu\text{mol kg}^{-1}$, which is approximately 45% of that now present at the surface of the Southern Ocean. This corresponds to an enhancement in the export flux of particulate organic matter

(POM) of about 135%. This is consistent with proxy evidence in the Southern Ocean, such as $\delta^{15}\text{N}$ in organic matter, interpreted as indicating increased nutrient utilization at the Last Glacial Maximum (LGM)^{24,25}. Although diatoms account for over 90% of the modelled increase in POM at the LGM, inferred opal accumulation rates in the real ocean suggest relatively little overall increase in opal export at this time²⁶. These apparently conflicting signals may be reconciled if there is more efficient use of silicate by diatoms as a result of increased Fe availability. Indeed, the model increase in opal export is only about 40%, less than one-third that exhibited by POM. Thus glacial-interglacial variability in the strength of biological surface silicon and carbon depletions is fundamentally decoupled by changes in Fe supply.

The amplitude of the atmospheric CO_2 change predicted by the model is limited to about 40 p.p.m. Of this, about 35 p.p.m. is through a ‘fast’ response of biological productivity to increased Fe availability, consistent with earlier models¹³. This is due to the lowering of Southern Ocean surface f_{CO_2} , which is about $60 \mu\text{atm}$ in the model, of the same order as may have finally been obtained in the SOIRE patch. About 5 p.p.m. of the model response occurs more gradually, as a lagged response of the system arising from “calcium carbonate compensation”¹⁵. By altering the parametrization of diatom cellular Si:C ratios, it is possible to make the dust signal drive more of the CO_2 rise at the terminations. However, Cd/Ca proxy evidence constrains productivity changes to be relatively modest²⁷. We therefore believe that it is most likely that one or more additional mechanisms to increase the concentration of atmospheric CO_2 must start once the termination is underway to account for the observed deglacial increase of around 80 p.p.m.v. (ref. 3).

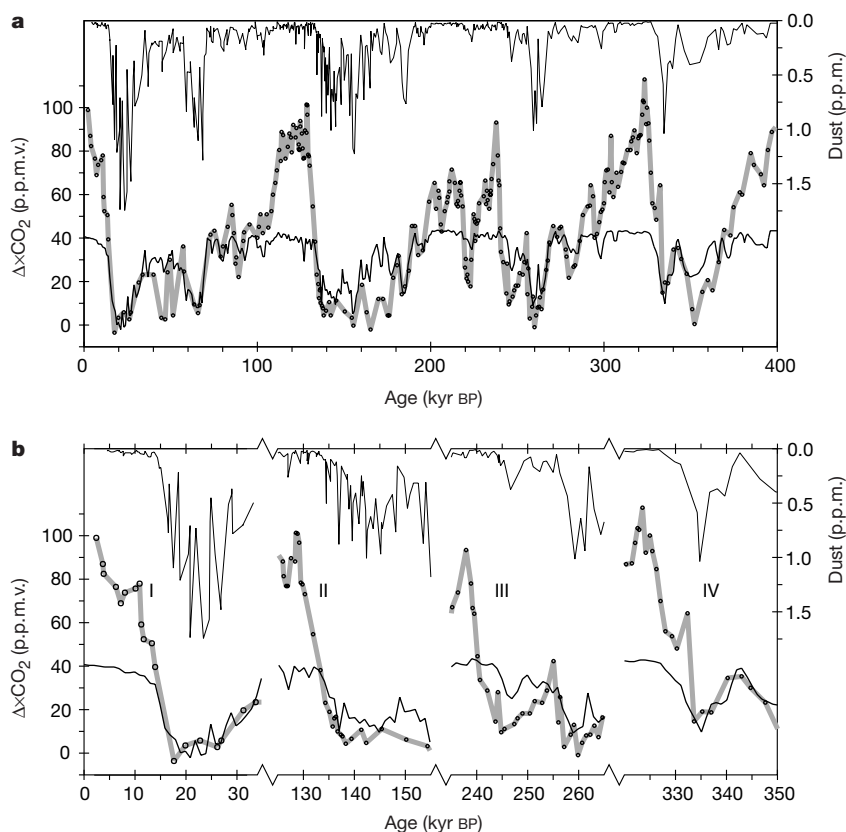


Figure 2 Vostok ice-core record³ of dust and atmospheric CO_2 compared to the atmospheric CO_2 concentration from our carbon-cycle model. **a**, The Vostok core dust record (inverted, at top), is used to force the model. The Vostok CO_2 record (thick grey line) is compared to the modelled atmospheric CO_2 (lower black line). The ice-core data are plotted using the age models described in ref. 3. The observed and modelled CO_2 mixing

ratios ($x\text{CO}_2$) are plotted as deviations from their LGM value ($\Delta x\text{CO}_2$). CO_2 levels in the model remain weakly sensitive to iron fluxes even when Fe is abundant in glacial times, owing to the dependence of Si:C ratios on Fe concentrations. **b**, Sections of the plot in **a** covering the four terminations (labelled I to IV) on a magnified timescale to show better the relative timing of the observed and modelled CO_2 records.

Possible candidates at the deglaciation include other Southern Ocean mechanisms, such as changes in sea-ice extent²⁸, mixing²⁹ and surface stratification^{24,25}.

Models of the processes influencing atmospheric CO₂ concentrations indicate that carbon might be sequestered from the atmosphere for long periods (centuries or longer) by deliberate fertilization of the Southern Ocean, though early suggestions that this would sequester carbon rapidly enough to substantially slow the increase of atmospheric CO₂ have been shown to be incorrect³⁰. Our observations confirm that stimulation of carbon fixation in the Southern Ocean is possible by adding iron. Integration of the deficit in dissolved inorganic CO₂ gives 1,200 (±10%) tonnes of excess C photosynthetically fixed in the SOIREE patch by day 12 of the experiment, larger than the independent estimate of extra algal carbon biomass⁴. However, our estimates of the amount removed from the atmosphere by the SOIREE bloom remain uncertain. No enhanced export flux was detected during the first 13 days of the experiment⁴. If the bloom that we induced eventually produced enhanced export, or the water at the surface was subducted on a timescale of a year or less, similar amounts of carbon could be sequestered from the atmosphere for long periods, though semi-continuous fertilization might be needed to achieve lasting sequestration³⁰. Otherwise, if there was no enhanced sinking flux, and the phytoplankton biomass was re-mineralized in the surface water and remained at the surface without subduction, carbon taken up from the atmosphere would be returned there on a timescale of a year or so. □

Methods

Ocean-atmosphere carbon-cycle box model

This model is based on a simplified representation of ocean circulation previously used in carbon-cycle investigations^{24,25,31}. Tracers advected in the ocean component include total dissolved inorganic carbon (ΣCO₂), dissolved oxygen (O₂), alkalinity, temperature and salinity. Of these, CO₂ and O₂ are exchanged with a 'well-mixed' atmosphere across the air-sea interface. The stable isotopes of carbon (¹²C and ¹³C) are treated separately, with all significant fractionation processes between them taken into account.

We describe the elements of the biological model briefly here, and in more detail in the Supplementary Information. Three nutrients potentially limiting to biological activity in the ocean are considered: phosphate (PO₄³⁻), silicic acid (H₄SiO₄), and total dissolved iron (Fe). The export of particulate organic matter (POM) from the surface ocean is estimated directly from ambient concentrations of these nutrients (and further modified to reflect temperature and insolation effects). This is a common simplification in global carbon-cycle models, circumventing the need for a full ecological model in the euphotic zone. We explicitly consider export arising from two distinct phytoplankton classes: diatoms and non-diatoms. Diatom productivity is assumed to be responsible for the export of opal, and is sensitive to the availability of H₄SiO₄. In contrast, non-diatom productivity is assumed to be responsible for the export of CaCO₃ (at a constant fraction), and is not directly influenced by ambient [H₄SiO₄]. Both classes contribute to export of POM, and as such share a common control by PO₄³⁻ and Fe availability.

Net Fe:C uptake ratios in both phytoplankton classes, and Si:C in diatoms, are varied as a function of [Fe]. In the former case, the dependence of this ratio on Fe availability is derived empirically from incubation results¹⁹. In the latter, we use changes in this ratio between Fe-replete and Fe-depleted conditions observed during this and previous studies^{10,11}. A continuous function for the Si:C ratio under intermediate degrees of Fe limitation is constructed, based on the assumption that, while the production rate of organic material is increasingly restricted with decreasing Fe availability, the rate of deposition of opal in diatoms is comparatively invariant (see Supplementary Information). The assumption is made in our model that changes in phytoplankton Fe:C and Si:C ratios are reflected proportionally in final export ratios. The composition of POM is characterized by a fixed ratio of C:N:P.

The biogeochemical cycling of dissolved iron in the ocean is treated in a similar manner to recent modelling work¹³. However, we assume that the scavenging rate of iron is proportional to the flux of particulate material falling through the water column, and do not account for the action of Fe-binding ligands. Particulate material exported from the euphotic zone is partially remineralized/dissolved within the water column. For simplicity, all POM and opal reaching the deep-sea sediment is returned in dissolved form to the ocean. An exception is made for Fe compounds (and complexes), which are removed from the system in order to balance their aeolian input at the ocean surface. The net accumulation of CaCO₃ in sediments is estimated as a function of deep-water chemical conditions, CaCO₃ and POM rain rates, and sedimentary CaCO₃ content³². The ocean-atmosphere system is therefore 'closed' with respect to PO₄ and H₄SiO₄ (and O₂), but 'open' with respect to CO₂, alkalinity and Fe.

In order to obtain the results shown in Fig. 2 the model was configured for the present-day carbon cycle, using distributions of dust deposition taken from general circulation model (GCM) simulations¹⁷. Fe is assumed to constitute 3.5% (by mass) of the dust flux,

of which 2% is soluble in the Southern Ocean. Biological productivity and Fe scavenging schemes were 'tuned' to give a reasonable match of model ocean distributions of [PO₄³⁻], [H₄SiO₄] and [Fe] to observations, together with realistic deep-sea particulate matter rain rates. After 'spin-up', the model was run for 400 kyr, with the dust deposition flux to the Southern Ocean varied according to the Vostok dust concentration signal³. The dust flux was scaled so that the maximum value at the LGM was 25 times that for the present day, consistent with model estimates for the Southern Ocean deposition¹⁷. No other model boundary conditions were changed.

Received 5 January; accepted 14 September 2000.

- Martin, J. H. Glacial-interglacial CO₂ change; the iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Kumar, N. *et al.* Increased biological productivity and export production in the glacial Southern Ocean. *Nature* **378**, 675–680 (1995).
- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–701 (2000).
- Watson, A. J., Liss, P. S. & Duce, R. A. Design of a small scale iron enrichment experiment. *Limnol. Oceanogr.* **36**, 1960–1965 (1991).
- Martin, J. H. *et al.* Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**, 123–129 (1994).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
- Cooper, D. J., Watson, A. J. & Ling, R. D. Variation of P_{CO₂} along a North Atlantic shipping route (UK to the Caribbean): A year of automated observations. *Mar. Chem.* **60**, 147–164 (1998).
- Abraham, E. R. *et al.* Importance of stirring in the development of an iron-fertilized phytoplankton bloom. *Nature* **407**, 727–730 (2000).
- Takeda, S. Influence of iron availability on nutrient consumption ratio of diatoms in oceanic waters. *Nature* **393**, 774–777 (1998).
- Hutchins, D. A. & Bruland, K. W. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**, 561–564 (1998).
- Duce, R. A. & Tindale, N. W. Atmospheric transport of iron and its deposition in the ocean. *Limnol. Oceanogr.* **36**, 1715–1726 (1991).
- Lefèvre, N. & Watson, A. J. Modelling the geochemical cycle of iron in the oceans and its impact on atmospheric carbon dioxide concentrations. *Glob. Biogeochem. Cycles* **13**, 727–736 (1999).
- Sohrin, Y. *et al.* The distribution of Fe in the Australian sector of the Southern Ocean. *Deep-Sea Res.* **47**, 55–84 (2000).
- Broecker, W. S. & Peng, T.-H. The role of CaCO₃ compensation in the glacial to interglacial atmospheric CO₂ change. *Glob. Biogeochem. Cycles* **1**, 15–29 (1987).
- Stevens, D. P. & Ivchenko, V. O. The zonal momentum balance in an eddy-resolving general-circulation model of the Southern Ocean. *Q. J. R. Meteorol. Soc.* **123**, 929–951 (1997).
- Mahowald, N. *et al.* Dust sources and deposition during the last glacial maximum and current climate: a comparison of model results with paleodata from ice cores and marine sediments. *J. Geophys. Res.* **104**, 15895–15916 (1999).
- Jickells, T. D. & Spokes, L. J. in *The Biogeochemistry of Iron in Seawater* (eds Turner, D. R. & Hunter, K. A.) (Wiley, New York, in the press).
- Sunda, W. G. & Huntsman, S. A. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.* **50**, 189–206 (1995).
- Levitov, S., Burgett, R. & Boyer, T. *World Ocean Atlas 1994 Vol. 3, Nutrients* (NOAA Atlas NESDIS, US Dept of Commerce, Washington DC, 1994).
- Sarmiento, J. L. & Toggweiler, J. R. A new model for the role of the oceans in determining atmospheric P_{CO₂}. *Nature* **308**, 621–624 (1984).
- Smith, H. J., Fischer, H., Wahlen, M., Mastroianni, D. & Deck, B. Dual modes of the carbon cycle since the Last Glacial Maximum. *Nature* **400**, 248–250 (1999).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO₂ changes. *Paleoceanography* **13**, 352–364 (1998).
- Francois, R. *et al.* Contribution of Southern Ocean surface-water stratification to low atmospheric CO₂ concentrations during the last glacial period. *Nature* **389**, 929–935 (1997).
- Sigman, D. M., Altabet, M. A., Francois, R., McCorkle, D. C. & Gaillard, J. F. The isotopic composition of diatom-bound nitrogen in Southern Ocean sediments. *Paleoceanography* **14**, 118–134 (1999).
- Mortlock, R. A. *et al.* Evidence for lower productivity in the Antarctic Ocean during the last glaciation. *Nature* **351**, 220–222 (1991).
- Rickaby, R. E. M. & Elderfield, H. Planktonic foraminiferal Cd/Ca: Paleonutrients or paleotemperature? *Paleoceanography* **14**, 293–303 (1999).
- Stephens, B. B. & Keeling, R. F. The influence of Antarctic sea ice on glacial-interglacial CO₂ variations. *Nature* **404**, 171–174 (2000).
- Toggweiler, J. R. Variation of atmospheric CO₂ by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
- Peng, T. H. & Broecker, W. S. Dynamic limitations on the Antarctic iron fertilization strategy. *Nature* **349**, 227–229 (1991).
- Keir, R. S. On the late Pleistocene ocean geochemistry and circulation. *Paleoceanography* **3**, 413–445 (1988).
- Archer, D. Modeling the calcite lysocline. *J. Geophys. Res.* **96**, 17037–17050 (1991).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank K. Tanneberger, U. Schuster, and the officers and crew of the RV *Tangaroa* for their support. This work was supported by the United Kingdom Natural Environment Research Council and the New Zealand National Institute for Water and Atmospheric Research. We acknowledge support from the CEFIC iron salts committee.

Correspondence and requests for materials should be addressed to A.J.W. (e-mail: a.watson@uea.ac.uk).

The possible subduction of continental material to depths greater than 200 km

Kai Ye, Bolin Cong & Danian Ye

Laboratory of Lithosphere Tectonic Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, PO Box 9825, China

Determining the depth to which continental lithosphere can be subducted into the mantle at convergent plate boundaries is of importance for understanding the long-term growth of supercontinents as well as the dynamic processes that shape such margins. Recent discoveries of coesite and diamond in regional ultrahigh-pressure (UHP) metamorphic rocks has demonstrated that continental material can be subducted to depths of at least 120 km (ref. 1), and subduction to depths of 150–300 km has been inferred from garnet peridotites in orogenic UHP belts based on several indirect observations^{2–5}. But continental subduction to such depths is difficult to trace directly in natural UHP metamorphic crustal rocks by conventional mineralogical and petrological methods because of extensive late-stage recrystallization and the lack of a suitable pressure indicator. It has been predicted from experimental work, however, that solid-state dissolution of pyroxene should occur in garnet at depths greater than 150 km (refs 6–8). Here we report the observation of high concentrations of clinopyroxene, rutile and apatite exsolutions in garnet within eclogites from Yangkou in the Sulu UHP metamorphic belt, China. We interpret these data as resulting from the high-pressure formation of pyroxene solid solutions in subducted continental material. Appropriate conditions for the Na₂O concentrations and octahedral silicon observed in these samples are met at depths greater than 200 km.

Supercontinents are generally thought to grow through subduction and collision of smaller continental blocks—as is happening in north India and southwest China, where the Indian continental plate is subducting below the Eurasian continent, and causing the uplift of the Tibetan plateau. Subduction of the South China continent under the North China continent in the Triassic⁹ period consolidated the old Eurasia continent, and resulted in the formation of various coesite- and diamond-bearing UHP metamorphic rocks in the orogen in eastern China (for a review, see ref. 10). The Sulu UHP metamorphic belt is the most deeply exhumed part of the Dabieshan-Sulu orogen. This provides the possibility of tracing the subduction depths of continental materials from fresh samples in the Sulu UHP belt.

Unusual eclogites were found as lenses in a garnet-peridotite layer intercalated with layers of UHP metamorphosed crustal igneous composite¹¹ at Yangkou. Partial transformation from low-pressure igneous textures (plagioclase stable) to UHP textures (coesite stable) have been reported in some meta-gabbroic layers of the composite¹². Unlike the common eclogites with near equigranular texture in the world-wide UHP belts, these eclogites contain an older generation of garnet (grt-I) which occurs as larger (up to 2 cm), purple, plastically deformed porphyroblast in the matrix of later medium-grained (0.05–0.5 mm) garnet (grt-II), omphacite (cpx-II), rutile (rut-II) and apatite (ap-II) (Fig. 1a). Grt-I exhibits abundant fine-grained euhedral rods or needles of clinopyroxene (cpx-I), rutile (rut-I) and apatite (ap-I) (Fig. 1a). The clinopyroxene rods are spindle-shaped in the long axis section (Fig. 1b), and hexagonal prismatic in the basal section (Fig. 1c). Apatite rods occur as hexagonal prism or rhombohedra. Like the hexagonal prismatic clinopyroxene, each group of apatite crystal planes is parallel to one group of rutile rods. In cases, several short apatite rods are dotted

along one long rutile rod (Fig. 1d). Backscattered electron images (see Supplementary Information) were used to estimate the amounts of the exsolutions. The rods increase in concentration with distance away from a rod-free edge zone (50–400 µm wide) to a maximum of 4–7 vol.% for clinopyroxene, 1–2 vol.% for rutile and 0.5–1 vol.% for apatite in the core.

Overall, the rods occur in groups of parallel rods; each group is orientated in one of four directions. The effect is to show an equilateral triangular form in cross-section (Fig. 1a). Universal-stage measurements show that the long-axis orientations of the rods are consistent with four cubic $\langle 111 \rangle$ directions. In a periodic cylindrical packing model¹³, these are dominant glide planes. Their concentration, topography and orientation suggest that the rods of clinopyroxene, rutile and apatite in the porphyroblastic garnet are generated by exsolution, rather than by epitaxial replacement or accidental inclusion. It is reasonable to assume that the cubic oxygen sublattice of garnet would have the same relation to the pseudohexagonal oxygen sublattice of clinopyroxene. This would imply a topotaxy between garnet and hexagonal prismatic cpx in Fig. 1c as: $(100)_{\text{cpx}} // (110)_{\text{grt}}; (010)_{\text{cpx}} // (\bar{1}10)_{\text{grt}}; (\bar{1}11)_{\text{cpx}} // (221)_{\text{grt}}$.

Mineral compositions were determined by means of a Camebax SX51 microprobe analyser. Longer counting times were programmed for Ti (20 s), Cr (20 s), P (60 s) and Na (60 s) in garnet than for other common elements (10 s). Representative analyses are listed in Table 1. The two stages of garnet and clinopyroxene differ widely in their composition. The porphyroblastic garnet shows a clear zonation. The exsolution-rich core contains higher Mg and Fe concentrations, and a lower Ca concentration. The exsolution-free rim has a composition similar to that of the matrix garnet. It contains higher Ca, and lower Fe and Mg (Table 1). Particularly high concentrations of P₂O₅ (0.06–0.17 wt%) and Na₂O (0.05–0.06 wt%) were detected in the core of the porphyroblastic garnet. However, the Ca-rich rim and the matrix garnet contain almost no P₂O₅ (< 0.03 wt%) or Na₂O (< 0.02 wt%) (Table 1). The exsolved cpx-I contains lower jadeite ($X_{\text{jd}} = 0.17–0.25$) than the matrix omphacite ($X_{\text{jd}} = 0.29–0.35$). (Here X_{jd} is the volume per cent of jadeite component.) The average composition of the exsolved cpx-I,

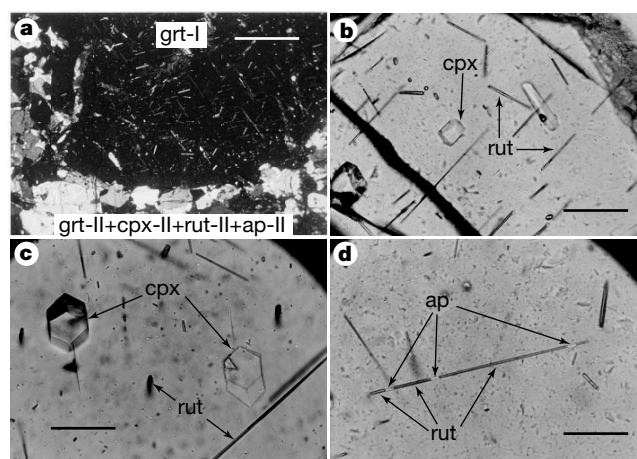


Figure 1 Transmitted light photomicrographs of garnet (grt) with exsolution rods of clinopyroxene (cpx), rutile (rut) and apatite (ap) along crystallographically controlled planes. **a**, Porphyroblastic garnet (grt-I) in the matrix of garnet (grt-II), omphacite (cpx-II), rutile (rut-II) and apatite (ap-II), the exsolution rods oriented in an equilateral triangular form. Scale bar, 1 mm. **b**, Spindle-shaped long-axis section of clinopyroxene rods in porphyroblastic garnet. Scale bar, 0.1 mm. **c**, Hexagonal prismatic basal sections of clinopyroxene rods in porphyroblastic garnet. Note that each group of pyroxene crystal planes is parallel to one group of needle-like rutile rods. Scale bar, 0.06 mm. **d**, Three short apatite rods dot one long rutile rod. Scale bar, 0.1 mm.

rut-I and ap-I were restored to the core composition of grt-I following the estimated volume proportions of $\text{grt}_{91.5}\text{cpx}_{7}\text{rut}_{1}\text{ap}_{0.5}$. The calculated original garnet (no. 6 in Table 1) is depleted in Al (with respect to the chemical formula) and contains 0.04 more Si than the ideal value of 3 for 12 oxygens, necessitating the existence of $0.06 (\text{Si}^{\text{IV}}_{\text{Total}} + \text{P}^{\text{IV}} - 3.00)$ 6-fold Si in octahedral coordination (Table 1). Here the roman numerals indicate the degree of coordination of the atoms. It also exhibits high Na_2O (0.34 wt%), TiO_2 (1.07 wt%) and P_2O_5 (0.28 wt%) values. A slightly lower average Na_2O content (0.28 wt%) was confirmed for the core of the porphyroblastic garnet with the highest amount of clinopyroxene exsolution by broad-beam electron microprobe analysis.

Experiments have predicted the structural transformation of pyroxene to garnet and the solid-state dissolution of pyroxene into garnet at very-high-pressure conditions (5–17 GPa)^{6–8}. Conversion of pyroxene to the garnet structure requires some of the four-fold tetrahedral Si in pyroxene to be changed to six-fold octahedral coordination in converted majoritic garnet, because the ratio of Si to O in pyroxene is 4:12, rather than the garnet value of 3:12. From a crystal chemistry point of view, substitution of $\text{Na}^{\text{VIII}} + \text{Si}^{\text{VI}} \rightarrow \text{Ca}^{\text{VIII}} + \text{Al}^{\text{VI}}$ in the garnet structure can describe these transformations¹⁴. Direct access to xenoliths sampled by diamond kimberlite pipes from the stable mantle lithosphere at depths greater than 300 km have revealed pyroxene dissolution in homogeneous garnet included in diamond¹⁵, and decompression resulted in exsolution of clinopyroxene in garnet¹⁶, which is identical to the occurrence of clinopyroxene in garnet reported here. Experiments^{17,18} have shown that garnet can host high Ti, Na and P contents under very-high-pressure conditions. Phosphorus has been suggested¹⁸ to enter the garnet structure through the coupled substitution of $\text{P}^{\text{IV}} + \text{Na}^{\text{VIII}} \rightarrow \text{Si}^{\text{IV}} + \text{Ca}^{\text{VIII}}$. Ringwood and co-workers^{6,17} proposed that $(\text{CaNa}_2)\text{Ti}_2\text{Si}_3\text{O}_{12}$ and $(\text{Ca}_2\text{Na})(\text{AlTi})\text{Si}_3\text{O}_{12}$ are probable Ti-bearing garnet components, based on compositions of garnets synthesized under high-pressure conditions, and they hence suggested that Ti enter garnet through substitution of $\text{Na}^{\text{VIII}} + \text{Ti}^{\text{VI}} \rightarrow \text{Ca}^{\text{VIII}} + \text{Al}^{\text{VI}}$. Such Na-, P- and Ti-bearing garnets had been reported in ultra-deep xenoliths in diamond kimberlite pipes^{14,15} and UHP metamorphic crustal rocks¹⁹.

Because of the lack of a suitable pressure indicator for the

bimineralic eclogitic assemblage, the pressure conditions under which the garnet was once homogeneously pyroxene-substituted must be inferred from experimental data on basaltic systems under subduction conditions. As pressure increases, the solubility of pyroxene in eclogitic garnet increases gradually^{7,20}, and the Na, Ti and P contents in garnet also have a positive pressure dependence¹⁸. The calculated Na_2O content (0.34 wt%) and octahedral silicon values (0.06) of the original Yangkou garnet are higher than those of the garnets synthesized²¹ at 7 GPa and 1,000 °C using starting material of a natural eclogite from Dabieshan, and also higher than those of garnet synthesized²⁰ at 7 GPa and 1,400 °C in a MORB system: but the values of 0.34 wt% and 0.06 are consistent with garnet synthesized⁷ at 7.6 GPa and 1,200 °C in an olivine tholeiite system. In these synthesis experiments, the temperatures of 1,400 °C and 1,200 °C are higher than in the subduction regime, and the Na_2O content in garnet is known²⁰ to be higher at higher temperatures. Thus, it is reasonable to conclude that the original garnet in the Yangkou eclogites was formed at pressures greater than 7 GPa. This puts the origin of the studied Yangkou eclogites at depths greater than 200 km.

The tectonic settings of the protolith of Yangkou eclogite and the host garnet-peridotite are not clear. However, lower-pressure mineral inclusions have been reported in garnet of similar garnet-peridotite from nearby localities in the same belt²². This implies possible crustal association of the studied eclogite before subduction. An alternative interpretation is that the peridotite was underplated at the base of the continental basement then underwent ultrahigh-pressure metamorphism²². But regardless of the exact tectonic settings, our results have implications for the dynamics of continental subduction and exhumations of the subducted rocks.

Density calculations based on experimental data confirm that subducted oceanic slab is substantially denser than the surrounding pyroclitic mantle at depths of 100–650 km (ref. 7). Oceanic slab has been observed to be subducted to depths exceeding 600 km in the mantle²³. Subduction of the less-dense continental crust to great mantle depths is traditionally thought to be difficult, because the operation of buoyancy forces would drive it back to the crustal level. However, subducted continental crust, and terrigenous and pelagic sediments, can be important in the geochemical evolution of the

Table 1 Representative compositions of clinopyroxene (cpx) and garnet (grt)

	1	2	3	4	5	6	7	8	9
Mineral	cpx	cpx	grt	grt	grt	grt	grt	grt	grt
Group	E	M	CP	RP	M	O	SY	SY	SY
SiO_2	55.36	55.56	39.39	38.93	39.21	40.32	41.43	41.49	40.94
TiO_2	0.06	0.01	0.11	0.06	0.06	1.07	0.15	0.03	1.49
Al_2O_3	6.14	7.90	21.37	21.01	20.87	20.18	22.46	22.77	22.56
Cr_2O_3	0.01	0.06	0.04	0.00	0.05	0.06	—	0.19	—
FeO	3.85	4.03	21.75	18.60	18.92	20.19	13.61	13.05	12.29
MgO	12.50	10.82	9.57	6.49	5.98	9.41	12.28	16.20	9.96
MnO	0.05	0.05	0.49	0.43	0.45	0.47	—	0.26	—
CaO	17.51	16.14	7.12	14.45	14.90	8.36	9.67	5.61	12.46
Na_2O	4.31	5.54	0.06	0.01	0.02	0.34	0.38	0.06	0.27
K_2O	0.00	0.01	0.00	0.00	0.00	0.00	—	0.06	—
P_2O_5	0.01	0.00	0.09	0.01	0.00	0.28	—	—	—
Total	99.78	100.12	99.99	99.99	100.46	100.68	99.98	99.72	99.97
Cation proportions based on 6 (for cpx) and 12 (for grt) oxygens									
Si	1.995	1.992	3.001	2.991	3.005	3.039	3.046	3.019	3.019
Ti	0.002	0.000	0.006	0.003	0.003	0.061	0.008	0.002	0.083
Al	0.261	0.334	1.919	1.902	1.885	1.793	1.946	1.953	1.961
Cr	0.000	0.002	0.002	0.000	0.003	0.004	—	0.011	—
Fe	0.116	0.121	1.386	1.195	1.213	1.273	0.837	0.794	0.758
Mg	0.671	0.578	1.087	0.743	0.683	1.057	1.346	1.757	1.095
Mn	0.001	0.002	0.032	0.028	0.029	0.030	—	0.016	—
Ca	0.676	0.620	0.581	1.189	1.224	0.675	0.762	0.437	0.984
Na	0.301	0.385	0.009	0.001	0.003	0.050	0.054	0.008	0.039
K	0.000	0.000	0.000	0.000	0.000	0.000	—	0.006	—
P	0.000	0.000	0.006	0.001	0.000	0.018	—	—	—

Values given are for analyses of minerals in Yangkou eclogites (nos 1–5); no. 6 represents calculated original garnet; nos 7, 8 and 9 are garnets synthesized in basaltic systems at 7.6 GPa and 1,200 °C by Irfune *et al.*⁷, 7 GPa and 1,000 °C by Okamoto and Maruyama²¹, and 7 GPa and 1,400 °C by Ono and Yasuda²⁰, respectively. Abbreviations: cpx: clinopyroxene; grt: garnet; E: exsolution; M: matrix; CP: core of porphyroblastic garnet; RP: rim of porphyroblastic garnet; O: original garnet; SY: synthesized garnet.

source of many alkali intraplate hotspot magmas—such sources are at extreme depths in the mantle^{24–26}. Recent discoveries of coesite and diamond in eclogites, metamorphic sedimentary rocks¹ and even granitic gneisses²⁷ in continental subduction orogenic belts reveal that large, coherent volumes of continental materials could be subducted to depths of more than 120 km. The present study further extends continental subduction to depths of more than 200 km. Experiments and density calculations⁸ show that the densities of the continental crust are substantially smaller than that of the surrounding pyrolitic mantle (3.4 g cm^{-3}) at pressures lower than 6 GPa (3.0 g cm^{-3}), slightly less-dense in the interval of 6.5–8.5 GPa ($3.2\text{--}3.3 \text{ g cm}^{-3}$), and much denser above 9–10 GPa until 24 GPa ($3.7\text{--}3.9 \text{ g cm}^{-3}$)⁸. Once the subducted continental materials reach a depth of about 200 km, their densities approach that of the surrounding mantle, and exceed the density of the surrounding mantle below 300 km. Thus, the buoyancy forces are effectively cancelled and entrainment by the sinking slab becomes the predominant influence. In the Sulu case, a subducted oceanic slab is thought to have hauled the following continental materials to depths greater than 200 km. Only one mechanism is reasonable for exhumation of such deeply subducted rocks: that they are rafted by rocks beneath them, propelled by buoyancy forces². Thus, a detachment was required along the subducting slab at some depths between 200 and 300 km, so that the buoyancy forces could drive the deeply subducted continental materials back to the crustal level.

Much larger-scale extremely deeply subducted crustal materials may be present in the Sulu UHP belt. Such materials would be indicated by the presence of appropriate mineralogical or micro-textural evidence—such as TiO_2 with the $\alpha\text{-PbO}_2$ structure (stable at 5–7 GPa), K-wadeite (6.5–7 GPa), or high-pressure C2/C clinoenstatite (6–7 GPa). We note that the presence of high-pressure C2/C clinoenstatite was previously predicted²⁸ for the Alpe Arami peridotite, and that this was subsequently confirmed³. □

Received 13 June; accepted 24 August 2000.

- Chopin, C. & Sobolev, N. V. in *Ultrahigh Pressure Metamorphism* (eds Coleman, R. G. & Wang, X.) 96–131 (Cambridge Univ. Press, New York, 1995).
- Dobrzhietskaya, L., Green, H. W. & Wang, S. Alpe Arami: a peridotite massif from depths of more than 300 kilometers. *Science* **271**, 1841–1845 (1996).
- Bozhilov, K. N., Green, H. W. & Dobrzhietskaya, L. Clinoenstatite in Alpe Arami peridotite: additional evidence of very high pressure. *Science* **284**, 128–132 (1999).
- Van Roermund, H. L. M. & Drury, M. R. An ultra-deep (>200 km) orogenic peridotite body in Western Norway. *Eos* **79**, F971 (1998).
- Yang, J. J., Godard, G., Kienast, J. R., Lu, Y. & Sun, J. Ultrahigh-pressure (60 kbar) magnesite-bearing garnet peridotites from northeastern Jiangsu, China. *J. Geol.* **101**, 541–554 (1993).
- Ringwood, A. E. & Major, A. Synthesis of majorite and other high pressure garnets and perovskites. *Earth Planet. Sci. Lett.* **12**, 411–418 (1971).
- Irfune, T., Sekine, T., Ringwood, A. E. & Hibberson, W. O. The eclogite-garnet transformation at high pressure and some geographical implications. *Earth Planet. Sci. Lett.* **77**, 245–256 (1986).
- Irfune, T., Ringwood, A. E. & Hibberson, W. O. Subduction of continental crust and terrigenous and pelagic sediments: an experimental study. *Earth Planet. Sci. Lett.* **126**, 351–386 (1994).
- Ames, L., Tilton, G. R. & Zhou, G. Z. Timing of collision of the Sino-Korean and Yangtze cratons: U-Pb zircon dating of coesite-bearing eclogites. *Geology* **21**, 339–342 (1993).
- Cong, B. L. *Ultrahigh-Pressure Metamorphic Rocks in the Dabieshan-Sulu Region of China* (Science Press/Kluwer Academic, Beijing, 1996).
- Wallis, S. R. et al. Occurrence and field relationships of ultrahigh-pressure metagranitoid and coesite eclogite in the Su-Lu terrane, eastern China. *J. Geol. Soc. Lond.* **154**, 45–54 (1997).
- Zhang, R. Y. & Liou, J. G. Partial transformation of gabbro to coesite-bearing eclogites from Yangkou, the Sulu terrane, eastern China. *J. Metamorph. Geol.* **15**, 183–202 (1997).
- Anderson, S. & O'Keeffe, M. Body-centered cubic cylinder packing and the garnet structure. *Nature* **267**, 605–606 (1977).
- Sobolev, N. V., Lavrent'ev, G. & Yu, G. Isomorphism sodium admixture in garnets formed at high pressure. *Contrib. Mineral. Petrol.* **31**, 1–12 (1971).
- Moore, R. O. & Gurney, J. J. Pyroxene solid solution in garnets included in diamond. *Nature* **318**, 553–555 (1985).
- Haggerty, S. E. & Sautter, V. Ultradeep (greater than 300 kilometers), ultramafic upper mantle xenoliths. *Science* **248**, 993–996 (1990).
- Ringwood, A. E. & Lovering, J. F. Significance of pyroxene-ilmenite intergrowths among kimberlite xenoliths. *Earth Planet. Sci. Lett.* **7**, 371–375 (1970).
- Tompson, R. N. Is upper-mantle phosphorus contained in sodic garnet? *Earth Planet. Sci. Lett.* **26**, 417–424 (1975).
- Schertl, H. P., Schreyer, W. & Chopin, C. The pyrope-coesite rocks and their country rocks at Parigi, Dora Maira Massif, Western Alps: detailed petrography, mineral chemistry and P-T path. *Contrib. Mineral. Petrol.* **108**, 1–21 (1991).

- Ono, S. & Yasuda, A. Compositional change of majoritic garnet in a MORB composition from 7 to 17 GPa and 1400 to 1600 °C. *Physics Earth Planet. Inter.* **96**, 171–179 (1996).
- Okamoto, K. & Maruyama, S. Multi-anvil re-equilibration experiments of a Dabie Shan ultrahigh-pressure eclogite within the diamond-stability fields. *Island Arc* **7**, 52–69 (1998).
- Zhang, R. Y., Liu, J. G. & Cong, B. Petrogenesis of garnet-bearing ultramafic rocks and associated eclogites in the Su-Lu ultrahigh-P metamorphic terrane, eastern China. *J. Metamorph. Geol.* **12**, 169–186 (1995).
- Widiyantoro, S. & Hist, R. Structure and evolution of lithospheric slab beneath the Sunda Arc, Indonesia. *Science* **271**, 1566–1570 (1996).
- Armstrong, R. L. Radiogenic isotopes: the case for crustal recycling on a near-steady state non-continental growth Earth. *Phil. Trans. R. Soc. Lond. A* **301**, 443–472 (1981).
- Dupre, B. & Allegre, C. Pb-Sr isotopic variations in Indian Ocean basalts and mixing phenomena. *Nature* **303**, 142–146 (1983).
- Loubet, M., Sassi, R. & Di Donato, G. Mantle heterogeneities: a combined isotope and trace element approach and evidence for recycled continental crust materials in some OIB sources. *Earth Planet. Sci. Lett.* **89**, 299–315 (1988).
- Ye, K. et al. Large areal extent of ultrahigh-pressure (UHP) metamorphism in the Sulu ultrahigh-pressure terrane of East China: new implications from coesite and omphacite inclusions in zircon of granitic gneiss. *Lithos* **52**, 157–164 (2000).
- Green, H. W., Dobrzhietskaya, L. & Bozhilov, K. N. Response to “Determining the origin of ultrahigh-pressure lherzolites. By Hacker, B. R., Sharp, T., Zhang, R. Y., Liou, J. G. & Hervig, R. L.” *Science* **278**, 704–707 (1997).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank P. Xu for programming microprobe analyses, and H. W. Green for comments that significantly improved this Letter. This work was supported by the National Natural Science Foundation of China

Correspondence and requests for materials should be addressed to K.Y. (e-mail: yekai@hotmail.com).

The population genetics of ecological specialization in evolving *Escherichia coli* populations

Vaughn S. Cooper & Richard E. Lenski

Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824, USA

When organisms adapt genetically to one environment, they may lose fitness in other environments^{1–4}. Two distinct population genetic processes can produce ecological specialization—mutation accumulation and antagonistic pleiotropy^{5–8}. In mutation accumulation, mutations become fixed by genetic drift in genes that are not maintained by selection; adaptation to one environment and loss of adaptation to another are caused by different mutations. Antagonistic pleiotropy arises from trade-offs, such that the same mutations that are beneficial in one environment are detrimental in another. In general, it is difficult to distinguish between these processes^{5–8}. We analysed the decay of unused catabolic functions in 12 lines of *Escherichia coli* propagated on glucose for 20,000 generations^{9,10}. During that time, several lines evolved high mutation rates¹¹. If mutation accumulation is important, their unused functions should decay more than the other lines, but no significant difference was observed. Moreover, most catabolic losses occurred early in the experiment when beneficial mutations were being rapidly fixed, a pattern predicted by antagonistic pleiotropy. Thus, antagonistic pleiotropy appears more important than mutation accumulation for the decay of unused catabolic functions in these populations.

We founded 12 populations from a strain of *E. coli*, and they all adapted to minimal medium supplemented with glucose^{9,10}. Competitive fitness increased rapidly in the first few thousand generations. Fitness continued to improve, but the average rate of

improvement decelerated sharply over time and was only about one-tenth as fast between generations 15,000 and 20,000 as during the first 5,000 generations (Fig. 1). While the evolving populations adapted to the glucose medium, unused catabolic functions decayed and their diet breadth became narrower and more specialized¹².

We sought to identify the population genetic process primarily responsible for the association between increased fitness in glucose and reduced diet breadth by following both characteristics over time. If most losses of catabolic function happened early in the experiment, when adaptation to glucose was most rapid, then antagonistic pleiotropy would be supported (Fig. 2, AP). Moreover, if the same functional losses occurred in most populations, then this parallelism would imply that the losses were adaptive and caused by antagonistic pleiotropy. Note that adaptation to glucose may result from mutations that either improve some aspect of glucose catabolism or eliminate unnecessary functions that are costly to fitness in glucose. In either case, the mutations improve fitness on glucose while adversely affecting performance on other substrates.

We do not expect this association between the dynamics of adaptation and decay under mutation accumulation. Rather, mutation accumulation predicts that losses of unused functions should accumulate stochastically, and different functions should decay in replicate populations. If one assumes that such losses occur by neutral mutations that knock out unused functions, then diet breadth should decay exponentially, that is log-linearly (Fig. 2, MA). 'Bottlenecks' caused by selective sweeps of beneficial alleles should not affect the expected rate of substitution of strictly neutral mutations, which depends on mutation rate but not on effective population size¹³. However, three of the 12 lines evolved defects in DNA mismatch repair, which led to genome-wide mutation rates about 50-fold higher than the rates experienced by the ancestor and other lines¹¹. These three lines became mutators around 2,500, 3,000 and 8,500 generations¹¹, and they retained high mutation rates through generation 20,000 (P. D. Sniegowski, unpublished data); therefore, all three experienced elevated mutation rates for most of the experiment. (A fourth line became a mutator between 16,500 and 18,000 generations (P. D. Sniegowski, unpublished data). Because most of this line's history occurred at the ancestral mutation rate, it was excluded from statistical comparisons between mutator and non-mutator populations at generation 20,000.) If mutation accumulation were the primary cause of ecological specialization, then the mutator lines should have evolved much narrower diet breadth than the lines that retained the ancestral mutation rate (Fig. 2, MA*).

We used Biolog ES plates to measure the catabolic function of the ancestor and three clones isolated from each population at generations 2,000, 10,000 and 20,000. The wells in these plates contain 95 different carbon sources and an indicator dye that reflects the amount of growth on the substrate; 64 substrates were informative

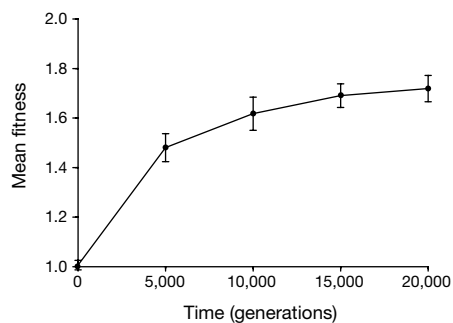


Figure 1 Trajectory for mean fitness of *E. coli* during 20,000 generations in minimal glucose medium. Each point is the mean of all 12 populations, and fitness of each population relative to the ancestor was measured with fivefold replication. Error bars are 95% confidence intervals based on the replicate populations.

for this study. On the basis of the proportion of genes in *E. coli* that encode these catabolic pathways, the plates can resolve mutations in several hundred genes (B. Bochner, personal communication). The weighted average of the 64 catabolic phenotypes provides an estimate of total catabolic function, or diet breadth.

Two predictions of antagonistic pleiotropy are that replicate lines should exhibit parallel decay of catabolic functions (those that trade-off with fitness in glucose), and that most losses should occur early (when adaptation to glucose is fastest). To test these predictions, for each substrate we compared the ancestral strain with the 12 evolved populations, as a group, at generations 2,000, 10,000 and 20,000. Because this approach entails multiple comparisons¹⁴, we were very conservative in our statistical criteria; we employed two-tailed *t*-tests, assumed unequal variances for the ancestor and evolved lines, and used 0.0005 as the critical *P*-value for hypothesis testing. A significant test indicates a parallel change in catabolic function that is common to most, if not all, of the lines (Fig. 3). Sixteen of the 64 informative substrates showed parallel decay in the first 10,000 generations. However, there was no further increase in this number over the next 10,000 generations, even though most of the informative substrates remained at risk (Fig. 3). In fact, nine substrates showed parallel decay after only 2,000 generations, at which time none of the populations had yet evolved a mutator phenotype. Average total catabolic function declined by 32% (from 1 to 0.68) in the first 10,000 generations, but declined only 15% (from 0.68 to 0.58) between 10,000 and 20,000 generations (Fig. 4). The initial decline was significant ($t = 9.712$, 11 d.f., one-tailed $P < 0.0001$), whereas the later decline was not ($t = 1.623$, 11 d.f., one-tailed $P = 0.0664$); the difference in proportional decay between the two periods was also significant (paired $t = 1.899$, 11 d.f., one-tailed $P = 0.0420$). The findings that there were many parallel reductions of catabolic function and that reductions were concentrated early in the experiment indicate that antagonistic pleiotropy contributed to ecological specialization.

This evidence does not exclude the possibility that mutation accumulation is also important, however. To that end, we tested whether the mutator populations accumulated more losses of catabolic function than did the populations that retained the ancestral mutation rate (Fig. 4). Although the observed trend was in the direction predicted by mutation accumulation, there was no significant difference in the catabolic diet breadth of the two groups after 20,000 generations ($t = 0.887$, 9 d.f., one-tailed $P = 0.1991$). Any difference between these groups in their extent of catabolic decay is small and subtle in comparison with the roughly 50-fold

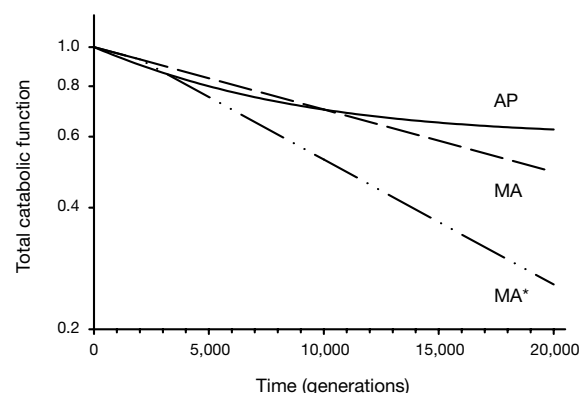


Figure 2 Hypothetical trajectories for the evolution of ecological specialization, as reflected by the decay of total catabolic function. AP, antagonistic pleiotropy, in which functional decay is inversely parallel to gains in fitness; MA, mutation accumulation (ancestral mutation rate), in which functional decay occurs at a constant rate and is independent of the pace of adaptation; MA*, mutation accumulation at an accelerated rate that occurs when a population becomes a mutator. Total catabolic function is shown on a log-transformed scale, such that predicted decay under mutation accumulation is linear.

Carbon source	Time (generations)		
	2,000	10,000	20,000
bromosuccinic acid	7	11	12
D-alanine	1	3	6
D-malic acid	5	12	12
D-ribose	12	12	12
D-saccharic acid	9	11	11
D-serine	12	11	10
D-sorbitol	12	11	11
fructose-6-phosphate	11	10	9
fumaric acid	9	12	12
glucose-1-phosphate	12	11	10
glucose-6-phosphate	11	12	8
glucuronamide	0	4	8
L-asparagine	8	12	12
L-aspartic acid	9	12	12
L-glutamine	12	12	12
L-lactic acid	11	12	10
L-malic acid	7	12	12
malic acid	9	12	12
mono-methylsuccinate	2	12	12
mucic acid	12	8	9
P-hydroxyphenylacetic acid	5	12	11
succinic acid	9	12	12
uridine	12	12	10
Sum of parallel losses	9	16	16

Figure 3 Summary of parallel changes in catabolic functions, based on comparisons between the evolved populations and common ancestor at three time points. Stringent significance ($P < 0.0005$) was demanded to account for multiple tests¹⁴ (64 substrates tested at each time point). Red, catabolic functions that consistently decayed; green, significant gains in function. The number in each cell is the number of populations (out of 12) whose average catabolic function on a substrate was less than that of the ancestor.

difference in mutation rates that they experienced for most of that time.

On balance, our data indicate that antagonistic pleiotropy was the main contributor to the resource specialization during evolution of *E. coli* lines in minimal glucose medium. That is, the mutations responsible for catabolic decay did not simply accumulate under mutation pressure but were themselves adaptive in the glucose environment. Two recent studies support this inference further. We have shown elsewhere that loss of the capacity to use D-ribose, which occurred in all 12 populations during the initial 2,000 generations, resulted from deletions that confer a selective advantage in glucose medium¹⁵. Funchain *et al.*¹⁶ studied *E. coli* mutator lines that evolved under a regime with severe population bottlenecks, which allowed deleterious mutations to accumulate. In contrast to our study, they found that (1) fitness declined over time; (2) losses of catabolic function identified using Biolog plates were not concentrated early in the experiment, but rather accumulated at a constant and perhaps even an accelerating rate; and (3) these defects were scattered across catabolic pathways, with no single defect being present in more than a small minority of their lines. The large population sizes and intense competition in our experiment would have prevented most deleterious mutations from being fixed as they would have been under a severe bottleneck regime.

We now consider three possible caveats concerning the hypothesized effects of mutation accumulation and antagonistic pleiotropy. First, we assumed that the amount of antagonistic pleiotropy was equivalent in all lines, which might be invalid if mutator populations experienced faster adaptive evolution than their less mutable counterparts. Consistent with our assumption, the three mutator lines did not attain significantly higher final fitness than the other eight lines ($t = 1.322$, 9 d.f., two-tailed $P = 0.2188$); nor does the mutation supply limit the rate of adaptive evolution under these circumstances¹⁷. Moreover, most parallel reductions in catabolic function occurred in the first 2,000 generations (Fig. 3), before any populations became mutators¹¹. Second, the prediction that unused functions should decay at a constant rate under mutation

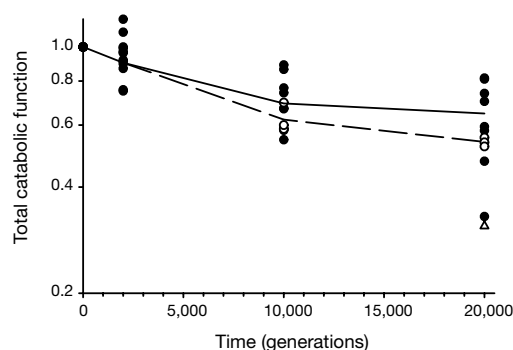


Figure 4 Evolution of total catabolic function during 20,000 generations in minimal glucose medium. Total catabolic function is a weighted average across 64 informative substrates; 1.0 is the ancestral level, while lower values indicate decay. Values are shown on a log-transformed scale, where the decay predicted by mutation accumulation is linear. Each point is the mean of three clones from each population. Solid line, mean value of populations that retained the ancestral mutation rate at that generation; dashed line, mean of three mutator populations; closed circles, populations with the ancestral mutation rate; open circles, mutator populations; open triangle, late appearing mutator (excluded from both classes).

accumulation holds true for neutral mutations¹³. But some of the decay might reflect deleterious mutations that hitch-hiked to fixation with beneficial mutations; during the early period of rapid adaptive evolution, there would have been more opportunities for deleterious mutations to hitch-hike, possibly generating a false indication of antagonistic pleiotropy. However, this explanation cannot account for the parallel decay of catabolic functions across replicate populations (Fig. 3). Also, we have demonstrated elsewhere that the mutations responsible for the decay of ribose catabolic function were indeed beneficial in the glucose medium¹⁵. Third, unused catabolic functions may not decay log-linearly under mutation accumulation if successive mutations in the same pathway interact epistatically¹⁸, or if some functions present larger mutational targets than others. These complications should have similarly affected the losses of catabolic function in the study by Funchain *et al.*¹⁶, which depended on mutation accumulation but not antagonistic pleiotropy. However, in that study, the rate of decay of catabolic functions tended to accelerate, in contrast to the deceleration observed here (Figs 3 and 4). Therefore, the deceleration in functional decay cannot be attributed to a more complex model of mutation accumulation, but instead indicates the important role of antagonistic pleiotropy.

In summary, a long-running evolutionary experiment with bacteria allowed us to distinguish between the effects of two processes that can contribute to ecological specialization, mutation accumulation and antagonistic pleiotropy. Our findings indicate that antagonistic pleiotropy played a greater role in specialization than did mutation accumulation. This conclusion does not contradict theory^{19,20} and experiments^{16,21,22} on the importance of mutation accumulation in populations that become less fit because of high mutation rates, small population sizes, or both. Rather, our work emphasizes the distinction between adaptive evolution, in which unused functions may be lost as a consequence of natural selection for other traits, and nonadaptive evolution, in which deleterious mutations accumulate precisely because selection is ineffective. □

Methods

Evolution experiment

The design of the long-term evolution experiment has been described in detail^{9,10}. In brief, 12 lines were founded from two variants of *E. coli* B and propagated daily by 1:100 dilution into Davis minimal medium supplemented with glucose at 25 $\mu\text{g ml}^{-1}$. The populations were maintained in this manner for 20,000 cell generations (3,000 days); every 500 generations, samples from each population were stored at -80°C . Six of the populations

were founded from a strain unable to grow on arabinose (Ara⁻), and the other six were founded with a spontaneous Ara⁺ mutant; the two ancestors were otherwise isogenic, and the Ara marker itself is neutral in the glucose medium⁹.

Fitness assays

The protocol for estimating the competitive fitness of evolved lines relative to their ancestor has been described⁸. In brief, samples of the evolved lines (containing whatever genetic diversity was present when they were sampled) and ancestral strains were removed from the freezer and separately acclimated to the medium and culture conditions used in the evolution experiment. Each evolved line was mixed with an equal volume of the reciprocally marked ancestor, and the two types then grew and competed under the same conditions that prevailed during the evolution experiment. Initial and final densities of the two competitors were enumerated by plating cells on a tetrazolium-arabinose indicator agar that allowed them to be distinguished by the Ara marker. (A different plating procedure was used for competitions with one evolved line that no longer produced distinct colonies on the indicator agar). The net growth rate of each competitor was calculated from the data, and the relative fitness of an evolved line is then expressed simply as the ratio of its growth rate to that of the ancestor. Assays were run in blocks with fivefold replication for all 12 lines.

Biolog assays

Catabolic diet breadth was assayed using Biolog (Hayward, California) ES plates for the two ancestral variants and three clones randomly chosen from each evolved line at generations 2,000, 10,000 and 20,000. Assays were run in three sets, each set comprising all of the clones for four lines plus three replicates of each ancestral variant. The bacteria were grown for two days in LB broth; on the next day, each culture was diluted 1:100 into fresh LB and incubated for 6 h. (LB was used instead of minimal glucose medium to avoid catabolite repression, which depresses other functions and may yield fewer positive readings²³.) The cultures were centrifuged at 12,000 g for 10 min and resuspended in saline to remove residual medium; this suspension was used to inoculate each well at a constant density. At 0, 4, 12, 24 and 48 h, optical densities were measured at 590 nm using an automated plate reader, and all measurements were adjusted by subtracting out the reading from the blank well. A trapezoidal area approximation²⁴ was used to integrate the five measurements for each well into one value, which reflects the area beneath the curve of optical density versus time; this area value is sensitive to both the rate and final level of catabolic function. Of the 95 substrates, glucose and arabinose were excluded *a priori* because glucose was the target of adaptation and arabinose use was a marker in the evolution experiment. Another 29 substrates were excluded because repeated measurements on the ancestor were statistically unreliable (coefficient of variation > 1), leaving 64 informative substrates. To test the evolutionary change in each individual catabolic function, the values for the three clones from a line at a given generation were averaged. The 12 evolved lines as a group were compared with the two ancestral variants using a two-tailed *t*-test with unequal variances (given divergence among the replicate lines) and a very stringent *P*-value of 0.0005 (to adjust for multiple tests¹⁴). Also, for each informative substrate, the catabolic function of an evolved clone was standardized to the ancestral value to give equal weight to all substrates, and then log-transformed to give equal weight to proportionally equivalent gains and losses of function. The anti-log of the average of these transformed values provides a measure of total catabolic function; the ancestral total equals 1.0 (by definition), whereas values less than 1.0 indicate an overall loss of function.

Received 28 March 2000; accepted 21 June 2000.

1. Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl Acad. Sci. USA* **58**, 217–224 (1967).
2. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19**, 207–233 (1988).
3. Fry, J. D. Tradeoffs in fitness on different hosts: evidence from a selection experiment with a phytophagous mite. *Am. Nat.* **136**, 569–580 (1990).
4. Bennett, A. F. & Lenski, R. E. Evolutionary adaptation to temperature. II. Thermal niches of experimental lines of *Escherichia coli*. *Evolution* **47**, 1–12 (1993).
5. Rose, M. R. & Charlesworth, B. A test of evolutionary theories of senescence. *Nature* **287**, 141–142 (1980).
6. Rose, M. R. *Evolutionary Biology of Aging* (Oxford Univ. Press, Oxford, 1991).
7. Holt, R. D. Demographic constraints in evolution: towards unifying the evolutionary theories of senescence and niche conservatism. *Evol. Ecol.* **10**, 1–11 (1996).
8. Sgrò, C. M. & Partridge, L. A delayed wave of death from reproduction in *Drosophila*. *Science* **286**, 2521–2524 (1999).
9. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).
10. Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814 (1994).
11. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* **387**, 703–705 (1997).
12. Cooper, V. S. Consequences of ecological specialization in experimental long-term evolving populations of *Escherichia coli*. Thesis, Michigan State Univ. (2000).
13. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
14. Miller, R. G. *Simultaneous Statistical Inference* (McGraw Hill, New York, 1981).
15. Cooper, V. S., Schneider, D., Blot, M. & Lenski, R. E. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *E. coli* B. *J. Bacteriol.* (submitted).
16. Funchain, P. *et al.* The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154**, 959–970 (2000).

17. De Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L. & Lenski, R. E. Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406 (1999).
18. Szathmáry, E. Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics* **133**, 127–132 (1993).
19. Muller, H. J. The relation of recombination to mutational advantage. *Mutat. Res.* **1**, 2–9 (1964).
20. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
21. Houle, D., Hoffmaster, D. K., Assimakopoulos, S. & Charlesworth, B. The genomic mutation rate for fitness in *Drosophila*. *Nature* **359**, 58–60 (1992).
22. Kibota, T. T. & Lynch, M. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**, 694–696 (1996).
23. *Biolog ES Microplate Instructions for Use* (Biolog, Hayward, California, 1993).
24. Guckert, J. B. *et al.* Community analysis by Biolog: curve integration for statistical analysis of activated sludge microbial habitats. *J. Microb. Meth.* **27**, 183–197 (1996).

Acknowledgements

We thank L. Ekunwe for assistance; J. Conner, J. Cooper, N. Cooper, D. Futuyma, D. Hall, A. Jarosz, T. Marsh, P. Moore, S. Remold and D. Rozen for discussions; and M. Blot, D. Schneider, P. Sniegowski and V. Souza for sharing unpublished data. This research was supported by NSF grants to V.S.C. and R.E.L. and by the Center for Microbial Ecology.

Correspondence and requests for materials should be addressed to V.S.C. (e-mail: cooperva@msu.edu).

Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*

Kenneth E. Filchak, Joseph B. Roethele & Jeffrey L. Feder

Department of Biological Sciences, Galvin Life Science Center, University of Notre Dame, Notre Dame, Indiana 46556-0169, USA

In *On the Origin of Species*, Darwin proposed that natural selection had a fundamental role in speciation¹. But this view receded during the Modern Synthesis when allopatric (geographic) models of speciation were integrated with genetic studies of hybrid sterility and inviability^{2,3}. The sympatric hypothesis posits that ecological specialization after a host shift can result in speciation in the absence of complete geographic isolation^{4,5}. The apple maggot, *Rhagoletis pomonella*, is a model for sympatric speciation in progress^{4,5}. Hawthorn (*Crataegus* spp.) is the native host for *R. pomonella* in N. America⁵. But in the mid-1800s, a new population formed on introduced, domesticated apple (*Malus pumila*)^{4,5}. Recent studies^{6–10} have conferred 'host race' status on apple flies as a potentially incipient species, partially isolated from haw flies owing to host-related adaptation. However, the source of selection that differentiates apple and haw flies is unresolved. Here we document a gene–environment interaction (fitness trade-off) that is related to host phenology and that genetically differentiates the races.

Because *Rhagoletis* flies mate exclusively on or near the fruit of their host plants¹¹, differences in host preference can result in virtually complete premating isolation among species¹². However, mark-release experiments have shown that such 'host fidelity' only partly restricts gene flow between apple and haw races to about 6% per year⁹ (*R. pomonella* is univoltine). Despite this exchange, the races are not fusing; allozyme loci mapping to three different regions of the fly's genome display consistent allele frequency differences over time¹⁰ (see Methods). Therefore, some form of strong host-dependent selection must be acting on these allozymes (or linked genes).

Evidence suggests that the interplay between host phenology, temperature and diapause is responsible for differentiating the

racers. Field data from Grant, Michigan (MI), USA, show that the earlier ripening time of sweeter apple varieties favourable for larval survival advances the seasonal distribution of apple flies by an average of 3–4 weeks before haw flies (Fig. 1). As a result, developing apple-fly larvae and pupae experience warmer temperatures and a longer time period before winter than haw flies. For example, daily internal fruit temperatures recorded at Grant in 1999 were on average 4 °C higher in apples (mean $\pm$ s.e., 24.1 ± 0.79 °C, 20–26 August) than haws (20.1 ± 0.46 °C, 13–17 September). (Note that these data probably underestimate conditions for larvae by 1–2 °C, as they were recorded during periods just after peak larval emergence from, rather than feeding within, fruits; Fig. 1.) In addition, *Rhagoletis* overwinters in a facultative pupal diapause; pre-imago flies will forgo a prolonged diapause and eclose within 30 days if

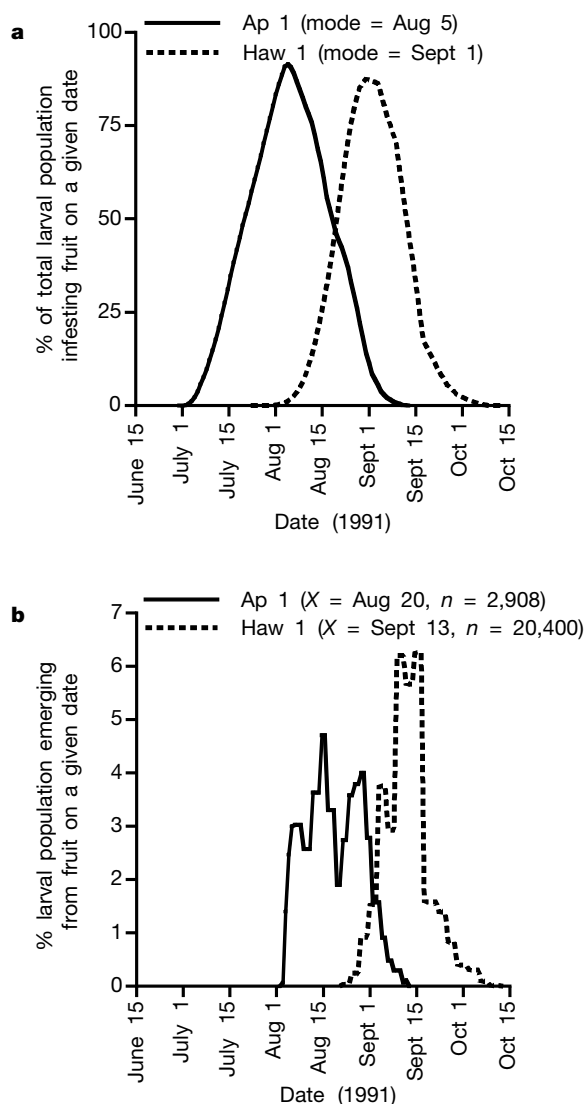


Figure 1 Seasonal distributions of fly larvae. Larvae feeding within (a), and emerging from (b), apple tree 1 & haw tree 1 fruit at Grant, MI, in 1991. (Note that these two trees support most of the fly population at Grant.) Details concerning the methods used to determine larval emergence/pupation dates can be found in experiments I–III of ref. 25. Larval feeding curves were generated by subtracting estimated development times for apple and haw flies (that is, the number of days from egg hatch in fruit to completion of larval feeding and pupation in the soil) from the distributions of larvae exiting apple and haw fruit in b. X is the mean. Methods for determining development rates in 1991 were similar to those for the ‘bag’ experiment in ref. 26. For a complete description of the *R. pomonella* life-cycle, see Supplementary Information.

exposed to elevated temperature¹⁰. In nature, such ‘non-diapausing’ flies are doomed because they either eclose as a second generation in the late fall when no fruit is present, or break diapause prematurely in the winter, depleting vital energy reserves and starving to death. Because allozyme alleles more common to apple flies at Grant, MI, all correlate with delayed adult eclosion and recalcitrance to non-diapause development¹⁰, we proposed that warmer conditions associated with the earlier phenology of apples select for a more muted developmental response (deeper diapause) in the apple than haw race.

To test our hypothesis, we reared haw-origin larvae and pupae under different combinations of prewinter temperature and winter length. We then scored eclosing adults for four allozymes displaying host-related frequency differences (*Dia-2*, *Me*, *Acon-2*, *Had*) and a control (*Idh*) that does not^{7,8}. Our prediction was that selection should favour ‘apple race’ alleles segregating in the haw-fly sample under the warmer prewinter temperature simulating apple (26 °C) and ‘haw race’ alleles under cooler, more haw-like conditions (17 and 22 °C). Longer winters were also expected to select against ‘haw’ alleles because pupae with these genes would tend to terminate diapause prematurely and die during extended chilling.

Results for the 26 °C treatment matched predictions. Except for *Dia-2*, frequencies of haw alleles all significantly fell with winter length (Fig. 2). These findings paralleled a previous study in which lengthening the 26 °C prewinter period also selected against haw genes¹⁰. In contrast, frequencies of haw alleles in the 17 and 22 °C treatments increased with winter length (Fig. 2). Genetic response curves therefore unexpectedly crossed between prewinter treatments, revealing a complex gene–environment interaction.

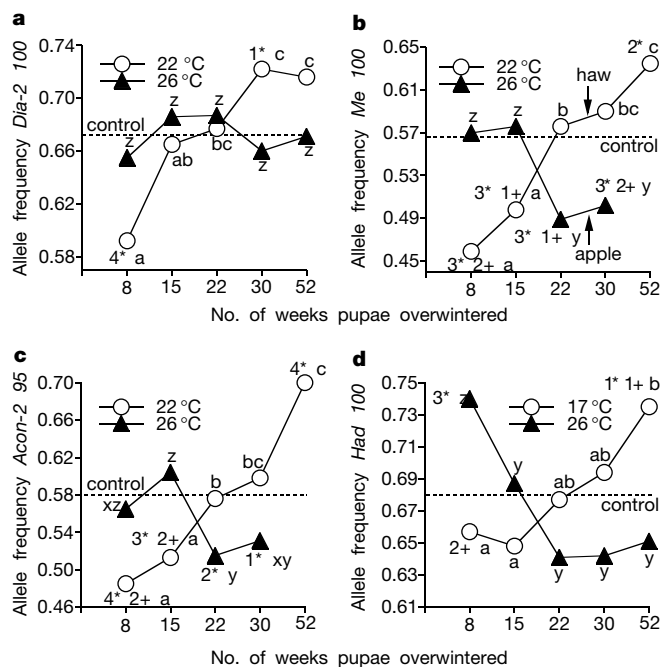


Figure 2 Comparisons of allele frequencies in eclosing adults. **a**, *Dia-2* 100; **b**, *Me* 100; and **c**, *Acon-2* 95 between 26 and 22 °C treatments. **d**, *Had* 100 between 26 and 17 °C treatments. Survivorship data, as well as genetic results for temperature treatments not shown and for the control locus *Idh*, are provided in Figs 2–4 of the Supplementary Information. Dashed lines are control allele frequencies. Asterisk indicates significant frequency difference between indicated sample and the control; ‘+’ indicates significant difference between prewinter treatments for a given winter length. Numerical prefix indicates significance level (1, $P \leq 0.05$; 2, $P \leq 0.01$; 3, $P \leq 0.001$; 4, $P \leq 0.0001$; Fisher exact tests). Samples within a prewinter treatment not sharing a common letter differ significantly in allele frequency. Arrows in **b** denote response at *Me* 100 for conditions best approximating those for apple and haw flies in nature.

Figure 3 depicts this interaction for *Me*, showing that absolute fitnesses for genotypes were negatively correlated across the 8- and 30-week winter periods within both the 26 °C ($r = -0.36$) and 22 °C ($r = -0.68$) treatments, indicative of fitness trade-offs. But the slopes of reaction norms had opposite signs, being negative with increased winter length for *Me 100* containing genotypes at 26 °C, and positive at 22 °C. Hence, the trade-offs result from a three-way interaction of gene, prewinter temperature and winter length, which implies that there is a threshold in the thermostatic regulation of diapause. Flies possessing haw race alleles that experience cooler temperatures (≤ 22 °C) before winter appear to set their diapause clocks to withstand long periods of chilling. However, warm temperatures (26 °C) cause these same genotypes to enter shallower diapauses, making them vulnerable to long winters. The control locus *Idh* did not respond to variation in either prewinter temperature or winter length.

The fitness trade-off can account for the genetic differentiation of the host races. Field data suggest that the 26 and 22 °C prewinter/30-week winter treatments approximate conditions faced by apple and haw flies, respectively, at our Michigan sites. (Soil temperatures at Grant in the winter of 1998–1999, as well as mean ambient recordings for Grand Rapids, MI, were below the *R. pomonella* developmental threshold¹³ of 8.7 °C for about 27 weeks from 15 October to 21 April.) Given the 30-week fitness values shown in Fig. 3, computer simulations indicate that *Me 100* frequencies would equilibrate at 0.18 in the apple and 0.64 in the haw race. The predicted difference of 0.46 between the races is greater than the actual difference of 0.20 ± 0.03 at Grant from 1984–1994. Comparable results were seen for predicted (*p*) and observed (*o*) differences for *Dia-2 100* ($p = 0.14$, $o = 0.14$), *Acon-2 95* ($p = 0.28$, $o = 0.26$) and *Had 100* ($p = 0.43$, $o = 0.04$). Details concerning how prewinter length and temperature (both daily fluctuations and means) interact to affect the genetic/physiological regulation of diapause are still needed to refine our model. We must also establish whether the allozymes themselves, or linked loci, are the targets of selection. However, allozyme surveys of natural populations indicate that the races are genetically tracking variation in temperature both temporally at sites and geographically with latitude, as predicted by our results¹⁰.

In conclusion, fitness trade-offs acting at the individual gene level are central to most models of sympatric speciation⁴. Unless the same alleles benefiting an insect on one plant are detrimental on others, gene flow and recombination could produce a 'jack-of-all-trades' genotype with high fitness on all hosts, eliminating the possibility of

sympatric divergence. Although reciprocal transplants have indicated that plants often act as different selective environments for insects^{14,15} and that host specialization can evolve quickly, it is unclear whether 'fundamental' trade-offs below the whole genome (organism) level were involved¹⁶. Here we show how divergent selection related to host phenology acts on host races of *R. pomonella*. Moreover, we demonstrate host-dependent trade-offs for the first time at the gene (genomic region) level, confirming a principal tenet of sympatric speciation. Notably, the observed allozyme response of the experimental haw population in both directions implies that considerable genetic variation exists within the races to respond to vagaries in local conditions, as well as to explore new plants with differing phenologies, potentially generating new host races and species. Other proposed cases of sympatric speciation in insects also appear to involve host/prey phenology^{17–19}, suggesting that host attributes affecting life-history timing may be as important as chemistry in influencing insect diet breadth and divergence. While we focused on the issue of sympatry, the message of a more direct role for natural selection and the environment in speciation applies regardless of geography. Indeed, conditions for ecological specialization are less restrictive in allopatry, due to relaxation of the constraint for fitness trade-offs^{20,21}. Moreover, the limited evidence for co-speciation between insects and plants²² suggests that adaptive radiation associated with host shifting is an important contributor to the great diversity of phytophagous insects. God's inordinate fondness for beetles may therefore reflect a more general fetish for phytophagous insects and their host plants. □

Methods

Fly collection and rearing

Larval-infested fruit were collected from a haw tree in E. Lansing, MI, on 5 September, 1997. Fruit were divided into three samples and held at 17, 22 or 26 °C in incubators (14/10 h light/dark cycle). Puparia were collected daily from each incubator and divided into six subsamples. One subsample was immediately frozen as a genetic control. The remaining pupae were put into petri-dishes containing moist vermiculite, returned to the incubators for 10 days, then kept at 14 °C for 2 days, before being placed in a refrigerator cycling between 0 and 5 °C to simulate winter. Dishes were removed from the cold after 8, 15, 22, 30 or 52 weeks, put in an incubator at 21 °C, and monitored for eclosing (surviving) adults over a 5-month period. Initial sample sizes were 1,352, 1,494 and 1,530 pupae for each winter length in the 17, 22 and 26 °C treatments, respectively.

Genetic analysis of flies

Standard starch gel techniques^{7,8} were used to score adults for NADH-diaphorase-2 (*Dia-2*), malic enzyme (*Me*), aconitase-2 (*Acon-2*) and hydroxyacid dehydrogenase (*Had*). These four allozymes display consistent allele frequency differences between apple (*a*) and haw (*h*) races at Grant, MI (mean frequencies 1984–1994: *Dia-2 100*, $a = 0.68 \pm 0.01$, $h = 0.82 \pm 0.01$; *Me 100*, $a = 0.42 \pm 0.02$, $h = 0.62 \pm 0.02$; *Acon-2 95*, $a = 0.20 \pm 0.01$, $h = 0.46 \pm 0.01$; *Had 100* $a = 0.79 \pm 0.01$, $h = 0.83 \pm 0.01$). *Dia-2* (chromosome 1), *Me* (linkage group 2), *Acon-2* (linkage group 2) and *Had* (linkage group 3) map to three different regions of the *R. pomonella* genome²³. We also scored adults for a control locus, isocitrate dehydrogenase (*Idh*), that does not show host-related differentiation^{7,8}. Numbers of adults scored for the 8–52-week winter treatments were 277, 310, 345, 377 and 184 for 17 °C; 205, 271, 311, 271 and 250 for 22 °C; and 334, 268, 195, 225 and 129 for 26 °C. We do not have data for *Me* and *Acon-2* for the 26 °C, 52-week sample because of technical difficulties. Results for untreated control samples were combined because no locus displayed significant heterogeneity among temperature treatments (total $n = 1,270$). We have sequenced an anonymous complementary DNA clone near *Dia-2*, *Me*, *Acon-2* and *Had* in linkage groups 1–3, and found two principal haplotypes segregating for each cDNA locus (our own unpublished data). These cDNA haplotypes were in strong gametic disequilibrium with the major 'apple' and 'haw' race electromorphs (alleles) segregating at the linked allozyme(s). The allozymes therefore reflect the allelic state of linked genes (genomic regions) and do not mask appreciable hidden genetic variation.

Temperature measurements

Internal fruit temperature was measured at Grant in 1999 using Hobo data loggers (catalogue no. H08-008-04; Onset Corp., Pocasset, MA). Thermal probes (TMC6-HC) were inserted into picked, ripe apples ($n = 4$, probe depth = 1.5 cm) and haws ($n = 4$, probe depth = 0.5 cm) placed in mottled sunlight, under the canopy on the west side of a haw tree, with temperature readings recorded automatically every 15 min. Winter length was estimated by measuring soil temperature at Grant from 1998–1999, as well as from ambient temperature data obtained over a 34-year period (1963–1996) from the national

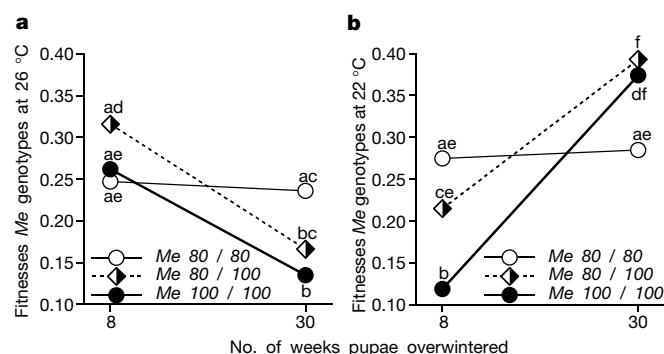


Figure 3 Absolute fitness estimates for *Me 100* genotypes related to winter length. **a**, 26 °C prewinter treatment; **b**, 22 °C prewinter treatment. Values were calculated by dividing raw genotype numbers surviving a given treatment by the expected numbers present at the start of the experiment based on frequencies in the untreated control. Estimates not sharing a common letter within or between the figures are significantly different, as determined from comparisons of 95% confidence intervals calculated assuming that survivorship follows a binomial distribution.

weather station at Grand Rapids, MI, located 35 km south of Grant. Soil measurements were made using Hobo data loggers (H08-031-08) fitted with probes (H08-031-08) buried 1.5 cm in the soil, the mean depth that *Rhagoletis* pupae overwinter²⁴.

Computer modelling of selection

Computer simulations were conducted using a discrete generation model mirroring the univoltine life-cycle of *Rhagoletis*. A random sample of 6% of the adult population eclosing under apple and haw trees was assumed to move to the alternate host each generation, matching gene flow estimates from mark-recapture studies⁹. Hard selection was invoked by weighting the proportions of immigrants and residents on a host by the mean fitness of the emigrant and resident race, respectively. After migration, adults randomly mated within host demes, with offspring experiencing viability selection based on absolute fitness values derived from the 26 °C (apple) or 22 °C (haw) prewinter, 30-week winter treatments. Given these values, the simulations converged on a stable equilibrium for each allozyme that maintained polymorphism and host-related differentiation regardless of initial allele frequencies.

Received 24 March; accepted 31 July 2000.

1. Darwin, C. *The Origin of Species by Means of Natural Selection* (Murray, London, 1859).
2. Dobzhansky, T. *Genetics and the Origins of Species* (Columbia Univ. Press, New York, 1937).
3. Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942).
4. Bush, G. L. in *Evolutionary Strategies of Parasitic Insects and Mites* (ed. Price, P. W.) 187–206 (Plenum, New York, 1975).
5. Bush, G. L. *The Taxonomy, Cytology and Evolution of the Genus Rhagoletis in North America* (Museum of Comparative Zoology, Cambridge, Massachusetts, 1966).
6. Prokopy, R. J., Diehl, S. R. & Cooley, S. S. Behavioral evidence for host races in *Rhagoletis pomonella* flies. *Oecologia* **76**, 138–147 (1988).
7. Feder, J. L., Chilcote, C. A. & Bush, G. L. Genetic differentiation between sympatric host races of *Rhagoletis pomonella*. *Nature* **336**, 61–64 (1988).
8. McPheron, B. A., Smith, D. C. & Berlocher, S. H. Genetic differences between *Rhagoletis pomonella* host races. *Nature* **336**, 64–66 (1988).
9. Feder, J. L. *et al.* Host fidelity is an effective pre-mating barrier between sympatric races of the apple maggot fly. *Proc. Natl Acad. Sci. USA* **91**, 7990–7994 (1994).
10. Feder, J. L. & Filchak, K. E. It's about time: The evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. *Ent. Exp. Appl.* **91**, 211–225 (1999).
11. Prokopy, R. J., Bennett, E. W. & Bush, G. L. Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). I. Site of assembly. *Canad. Ent.* **103**, 1405–1409 (1971).
12. Feder, J. L. & Bush, G. L. A field test of differential host plant usage between two sibling species *Rhagoletis pomonella* fruit flies (Diptera: Tephritidae) and its consequences for sympatric models of speciation. *Evolution* **43**, 1813–1819 (1989).
13. Reid, J. A. & Laing, J. E. Development threshold and degree-days to adult emergence for overwintering pupae of the apple maggot, *Rhagoletis pomonella* (Walsh) collected in Ontario. *Proc. Ent. Soc. Ontario* **197**, 19–22 (1976).
14. Via, S. in *Ecological Genetics* (ed. Real, L. A.) 58–85 (Princeton Univ. Press, Princeton, New Jersey, 1994).
15. Craig, T. P., Horner, J. D. & Itami, J. K. Hybridization studies on the host races of *Eurosta solidaginis*: Implications for sympatric speciation. *Evolution* **51**, 1552–1560 (1997).
16. Thompson, J. N. *The Coevolutionary Process* (Univ. Chicago Press, Chicago, 1994).
17. Abrahamson, W. G. *et al.* in *Gall-Forming Insects* (eds Price, P., Mattson, W. & Baranchikov, Y.) 208–222 (USDA Forest Service Tech. Report NC-174, St. Paul, Minnesota, 1994).
18. Wood, T. K. & Keese, M. C. Host-plant-induced assortative mating in *Echenopa* treehoppers. *Evolution* **44**, 619–628 (1990).
19. Tauber, C. A. & Tauber, M. J. in *Speciation and Its Consequences* (eds Otte, D. & Endler, J. A.) 307–344 (Sinauer, Sunderland, Massachusetts, 1989).
20. Fry, J. D. The evolution of host specialization: are trade-offs overrated? *Am. Nat.* **148**, S84–S107 (1996).
21. Feder, J. L. in *Endless Forms: Species and Speciation* (eds Howard, D. & Berlocher, S. H.) 130–144 (Oxford Univ. Press, New York, 1998).
22. Farrell, B. D. “Inordinate fondness” explained: Why are there so many beetles? *Science* **281**, 555–558 (1998).
23. Roethel, J. B. *et al.* Towards a molecular genetic linkage map for the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae): a comparison of alternative strategies. *Ann. Entomol. Soc. Am.* **90**, 470–479 (1997).
24. Lathrop, F. H. & Nickels, C. B. *The biology and control of the blueberry maggot in Washington County Maine* (Tech. Bull. 275 US Dept of Agriculture, Washington DC, 1932).
25. Feder, J. L. The effects of parasitoids on sympatric host races of the apple Maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae). *Ecology* **76**, 801–813 (1995).
26. Filchak, K. E. *et al.* A field test for host-plant dependent selection on larvae of the apple maggot fly, *Rhagoletis pomonella*. *Evolution* **53**, 187–220 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Berlocher, J. Carbol, H. Dambroski, G. Dwyer, V. Gavrilovic, D. Gibbons, D. Lodge, B. Perry, D. Prokrym, U. Stolz, S. Via and the USDA lab at Niles, Michigan. This work was supported by grants from NSF (Young investigators, Integrative Research Challenges & GRT) and the USDA to J.L.F. and the University of Notre Dame.

Correspondence and requests for materials should be addressed to K.E.F. (e-mail: Filchak.1@nd.edu).

Learning of action through adaptive combination of motor primitives

Kurt A. Thoroughman*† & Reza Shadmehr*

* Department of Biomedical Engineering, Johns Hopkins University, 419 Traylor Building, 720 Rutland Avenue, Baltimore, Maryland 21205, USA

Understanding how the brain constructs movements remains a fundamental challenge in neuroscience. The brain may control complex movements through flexible combination of motor primitives¹, where each primitive is an element of computation in the sensorimotor map that transforms desired limb trajectories into motor commands. Theoretical studies have shown that a system's ability to learn action depends on the shape of its primitives². Using a time-series analysis of error patterns, here we show that humans learn the dynamics of reaching movements through a flexible combination of primitives that have gaussian-like tuning functions encoding hand velocity. The wide tuning of the inferred primitives predicts limitations on the brain's ability to represent viscous dynamics. We find close agreement between the predicted limitations and the subjects' adaptation to new force fields. The mathematical properties of the derived primitives resemble the tuning curves of Purkinje cells in the cerebellum. The activity of these cells may encode primitives that underlie the learning of dynamics.

Studies of reaching movements have demonstrated that humans construct motor commands based on a prediction of forces that will be experienced in the upcoming movement³. When new forces are imposed on the arm, the prediction is in error and the arm does not follow the desired trajectory^{3,4}. With practice the motor commands are modified⁵ and the trajectory approximates the desired path. The learning of dynamics, however, affects movements outside the region of training^{3,6–8}, suggesting that the brain builds a state-dependent approximation of external forces⁹, called an internal model. Occasional movements with unexpectedly altered dynamics, termed ‘catch trials’, have been used to quantify how the internal model generalizes^{3,4}. Catch trials, however, not only test the internal model for a given movement but cause errors that in turn change the internal model and affect future movements. We demonstrate that the effect of errors experienced in a given movement on subsequent movements can reveal characteristics of primitives with which motor commands are generated.

We consider the internal model to be a sensorimotor map transforming desired arm trajectories into muscle forces^{10–12} through a flexible combination of a set of primitives:

$$\hat{\mathbf{f}} = \mathbf{W}^T \mathbf{g}(\mathbf{x}^*, \dot{\mathbf{x}}^*, \ddot{\mathbf{x}}^*) \quad (1)$$

where T is the transpose operator, $\hat{\mathbf{f}}$ is a vector approximation of forces $\mathbf{f}$ to be produced by muscles to compensate for task dynamics, and $\mathbf{g}$ is a vector of scalar-valued primitives $[g_1, \dots, g_J]^T$. Although in general $\mathbf{g}$ can depend on desired position, velocity and acceleration $(\mathbf{x}^*, \dot{\mathbf{x}}^*, \ddot{\mathbf{x}}^*)$, here we investigated learning of viscous forces and therefore considered a simpler subset of primitive functions that depended only on desired velocity. The internal model is learned through experience-dependent modification of the weight matrix $\mathbf{W}$. Assuming a learning rule that minimizes $\tilde{\mathbf{f}}^2 \equiv \|\mathbf{f} - \hat{\mathbf{f}}\|^2$, $\mathbf{W}$ is adjusted after a movement (indexed 1) according to:

$$\Delta \mathbf{W}_1 = -\eta \mathbf{g}(\dot{\mathbf{x}}_1^*) \tilde{\mathbf{f}}^T \quad (2)$$

where η is a constant learning step. This adaptation changes the

† Present address: Vollen Center of Complex Systems and Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA.

internal model output in the subsequent movement (indexed 2):

$$\begin{aligned}\hat{\mathbf{f}}_2(W + \Delta W_1) - \hat{\mathbf{f}}_2(W) &= (W + \Delta W_1)^T \mathbf{g}(\dot{\mathbf{x}}_2^*) - W^T \mathbf{g}(\dot{\mathbf{x}}_2^*) \\ &= -\eta \mathbf{g}^T(\dot{\mathbf{x}}_1^*) \mathbf{g}(\dot{\mathbf{x}}_2^*) \tilde{\mathbf{f}}_1\end{aligned}\quad (3)$$

The change in the internal model output depends on experienced error and the mutual projection between evaluations of the primitives, but does not depend on the weight matrix. As the primitives depend on desired velocity, when the two movements have the same desired trajectory (for example, toward the same target), the change should be proportional to the error experienced. When the two movements are toward different targets, the change will also depend upon the breadth of the receptive fields of the primitives.

We first tested whether an error experienced in a given movement causes a proportional change in the internal model for the next movement to the same target. We asked subjects to make reaching movements while holding a manipulandum¹³ which produced viscous forces $\mathbf{f} = B\dot{\mathbf{x}}$, where $B = \{0.13; -13.0\}$ N s m⁻¹. Catch trials, movements during which $\mathbf{f} = 0$, were randomly interspersed among the targets. Our proxy for error was hand displacement perpendicular to target direction (perpendicular displacement; p.d.) measured 250 ms into the movement. The first movement in the field (1st in Fig. 1a) had significant error (p.d. = 2.38 cm), but with training (ct-1 in Fig. 1a) became less disturbed (p.d. = 0.45 cm). In the next movement towards this direction (90°), a catch trial (ct), there was a large error in the direction opposite the initial error, suggesting formation of an internal model¹³. In the subsequent movement to 90° (ct+1), during which the force field was present, the p.d. was substantially greater (p.d. = 1.22 cm) than in ct-1, indicating partial unlearning of the internal model as predicted by equation (2). In agreement with equation (3), there

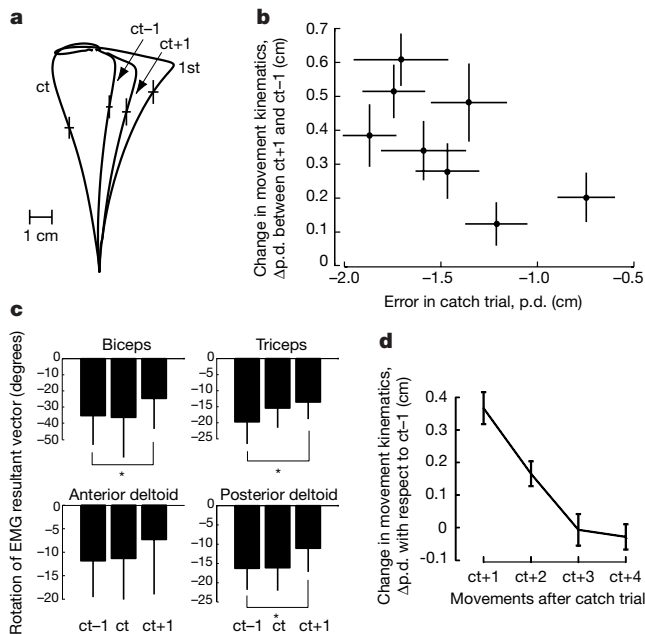


Figure 1 Catch trials induced short-term unlearning. **a**, Hand trajectories of single movements, averaged across all 40 subjects, during the first (labelled 1st), 38th (ct-1), 39th (ct) and 40th (ct+1) movements toward 90° (0° is at 3:00). Intersection of error bars indicates parallel and perpendicular displacement (p.d.) 250 ms into the movement. **b**, Jumps in p.d. between before (ct-1) and after (ct+1) catch trials versus p.d. in catch trials (ct), averaged within target directions across subjects. **c**, Orientation of preferred direction of movement-initiating EMG in ct-1, ct and ct+1. Asterisks indicate significant ($P < 0.05$) rotation of EMG preferred direction between ct-1 and ct+1. **d**, Errors in force field movements following ct. In all figures, error bars indicate 95% confidence intervals of the mean (CIM) across subjects.

was a significant correlation between magnitude of movement errors in the catch trial and the unlearning observed in ct+1 (Fig. 1b, $r = 0.65$). The physiological correlate of this unlearning was evident in the spatial tuning of movement-initiating muscle activations. The computational construct of an internal model predicted that spatial tuning of the electromyographic (EMG) activity of arm muscles would undergo a specific rotation with training⁵. During training, the preferred direction of this tuning gradually rotated. However, between ct-1 and ct+1 the preferred direction rotated back toward the initial orientation (Fig. 1c), indicating unlearning of the internal model. This unlearning was washed out by movement ct+3 (Fig. 1d).

We next investigated the shape of primitives underlying internal model formation by quantifying, independent of the model in equation (3), the temporal dynamics of movement errors first within and then across directions. In a sequence of random target directions, the time series of movement errors for a given direction was fitted to the following system of equations:

$$\begin{aligned}z_{n+1} &= az_n + bu_n \\ y_n &= z_n + du_n\end{aligned}\quad (4)$$

Here y represented error in the internal model as quantified by p.d., n was movement number and u indicated whether the force field

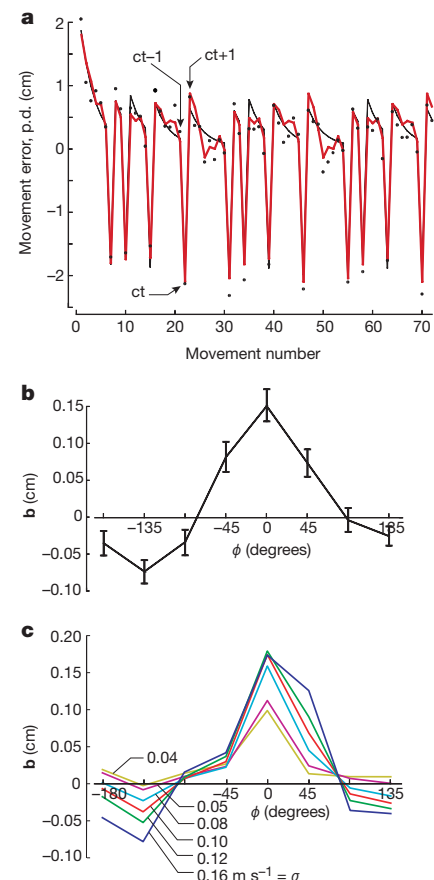


Figure 2 Sensitivity to movement error across target directions. **a**, Errors in subjects' reaches (circles) (p.d. at 250 ms) toward 90°, averaged across subjects ($n = 40$). Catch trials are the large negative spikes (one is labelled ct). Lines are best fits of scalar (black) and vector (red) state-space models to the data (y in equation (4)). **b**, Average sensitivity of the internal model to errors experienced in previous movements (b in equation (4)) as a function of angular distance ϕ . Error bars (95% CIM) calculated through bootstrapping. **c**, Average generalization function $b(\phi)$ for simulated adaptive controllers that constructed an internal model with Gaussians of width σ . $b(\phi)$ in simulations and subjects were averaged across target directions.

was present ($u = -1$) or turned off ($u = 1$). The hidden state of the system, z , represented the amount of movement error generated by the internal model; actual error (y) also depended on whether the force field was applied. The implicit assumption in this initial model was that errors experienced in one target direction did not affect the internal model for generating movements toward other targets. The best-fit model correlated to actual errors reasonably well (Fig. 2a, black line; across directions, mean $r = 0.60$). The fit mimicked subjects' recovery from initial error, their large error in the catch trial, and their jump in error from $ct-1$ to $ct+1$. Whereas this initial model smoothly decayed after jumps in error, subjects often generated a non-monotonic change in error between catch trials. We hypothesized that this was because errors experienced in one target direction changed the internal model for other directions.

To investigate whether locally experienced errors affected other directions of movement, we expanded both u and b in equation (4) to eight-dimensional vectors. Each element of the input vector u flagged recently experienced dynamics in a particular target

direction i : whether, since the last movement in the modelled target direction, a force-field movement ($u(i) = -1$) or a catch trial ($u(i) = 1$) had been most recently experienced, or if no movement had occurred in direction i ($u(i) = 0$). Each element of b , denoted $b(i)$, quantified the sensitivity to errors experienced in direction i . The expanded model now accounted for subtle changes in actual movement error (Fig. 2a, red line; mean $r = 0.81$). Confidence intervals on b suggested that there was a significant, nonzero influence of local errors on subsequent control in other directions. To calculate sensitivity across target directions, the elements of b were re-indexed by the angular distance between the direction in which errors were experienced and the modelled movement direction. This angular distance was represented by ϕ . Averaging $b(\phi)$ across movement directions (Fig. 2b) demonstrated that errors experienced in a given movement maximally influenced the internal model for that direction. This influence decayed in neighbouring directions. Surprisingly, sensitivity became significantly negative when angular distances were larger than 90° . This

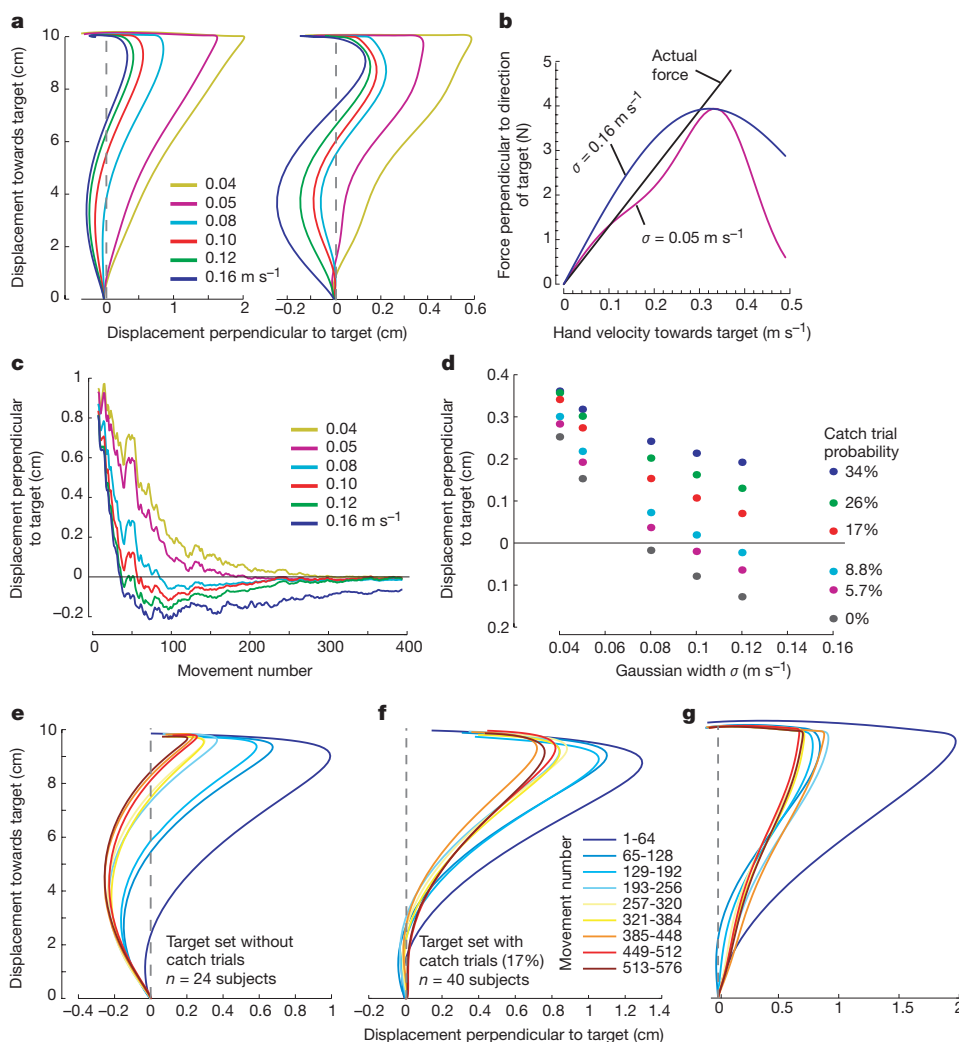


Figure 3 Movement characteristics of systems that learn an internal model with velocity encoding gaussians. **a**, Simulated adaptive controllers trained in a target set without catch trials. The 11th (left) and 21st (right) movements toward 90° are shown for each controller. Learning with wide gaussians produced S-shaped movements. **b**, Adaptive controllers' estimation of the force field after training for 100 movements in a target set without catch trials. Peak velocity of typical movements is 0.35 m s^{-1} . **c**, Time course of error (p.d. at 200 ms) for simulated controllers with various gaussian widths, smoothed across a 13-movement window. Overcompensation (generation of S-shaped hand trajectories) occurs when p.d. becomes negative. **d**, Error (p.d.) averaged over the subset

of movements 65–128 during which the force field was on. With narrow gaussians, overcompensation never occurred, regardless of catch trial probability. With wide gaussians, movements became S-shaped below critical probabilities (near 17%). **e, f**, Parallel and perpendicular displacements averaged across movements and subjects. **e**, Subjects ($n = 24$) trained in target sets without catch trials, resulting in S-shaped movements. **f**, Subjects ($n = 40$) trained in target sets with 17% catch trial probability and did not produce S-shaped movements. **g**, A controller relying upon wide gaussians ($\sigma = 0.12 \text{ m s}^{-1}$) trained in target sets with 17% catch trial probability did not produce S-shaped trajectories.

indicated that when two force-field movements were separated by angular distance ϕ , if ϕ was small, then errors experienced in the first movement improved the internal model for the second movement. If ϕ was large, then errors in the first movement destructively interfered with the internal model used to generate the second movement.

To explain this result, we note that sensitivity of the internal model to experienced errors, $\mathbf{b}(\phi)$, was quantified in terms of the p.d. Both the output and the error signal of the internal model, however, are in terms of force (equations (1) and (2)). Because the force field is linear in velocity, the direction of force error corresponding to positive (clockwise) p.d. towards one target opposes the direction of force error corresponding to a positive p.d. towards the opposite target. Interpreting the sensitivity of subjects' internal model (Fig. 2b) through the adaptation rule (equation (3)) suggests that both the positive values of $\mathbf{b}$ for $-45^\circ < \phi < 45^\circ$ and the negative values of $\mathbf{b}$ for large ϕ correspond to the same direction of force compensation. From this we deduced that the mutual projection $\mathbf{g}^T(\dot{\mathbf{x}}_i)\mathbf{g}^T(\dot{\mathbf{x}}_{i+1})$ declines but always remains positive as the angular distance between two movements increases. This result rules out bases that encode velocity space linearly. Furthermore, because information experienced in each direction most strongly affects that direction and its neighbour less so, basis functions that have specific regions of preferred activity are more likely to underlie learning than global representations of dynamics.

We therefore investigated what conditions on $\mathbf{g}(\dot{\mathbf{x}}^*)$ were sufficient to generate the generalization function $\mathbf{b}(\phi)$. A salient property of cells in the motor system is their directional tuning¹⁴ and modulation with hand speed¹⁵. In the cerebellum, a region which

lesion^{16–19} and functional imaging studies^{20,21} have linked to learning and control of arm dynamics, many Purkinje cells simultaneously encode the direction and speed components of velocity²². These cells broadly encode hand velocity during planar reaching, firing maximally at preferred velocities distributed in velocity space. This encoding precedes in time the actual movement, suggesting that these cells encode desired velocity. The behaviour of each cell k could therefore be represented as a gaussian with a centre located at position $\mathbf{c}_k$ in desired velocity space. We simulated a controller attached to a biomechanical model of the arm that learned an internal model with basis functions:

$$g_k(\dot{\mathbf{x}}^*) = \exp(-|\dot{\mathbf{x}}^* - \mathbf{c}_k|^2/2\sigma^2) \quad (5)$$

where σ is the standard deviation of the gaussian. To accommodate the possibility that the exact shape of $\mathbf{b}(\phi)$ depended on the training paradigm, we trained subjects and the simulated controller with the identical set of targets and catch trials. A crucial component of the simulations was σ , the width of the primitives. When the gaussians were narrow, the time series of errors generated by the simulation showed a spike after each catch trial and a smooth decay afterwards, similar to the scalar-input state-space model fit (Fig. 2a, black line) but unlike the performance of our subjects. Simulations driven by broad gaussians, however, produced non-monotonic changes that mimicked subjects' actual patterns of adaptation. Simulation results were fitted with equation (4) to produce the generalization function $\mathbf{b}(\phi)$ (Fig. 2c). The $\mathbf{b}(\phi)$ generated with narrow gaussians rapidly dropped to zero as ϕ changed from zero. Learning with wide gaussians, however, showed a generalization that was very similar to actual subject performance, including negative sensitivity for large

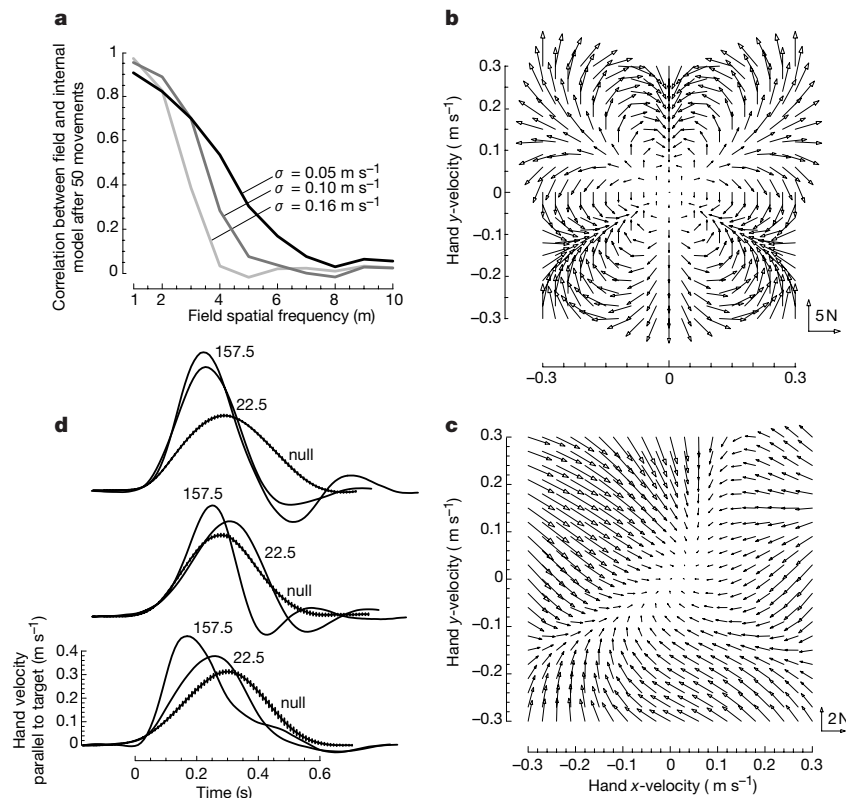


Figure 4 Learning with wide gaussians imposes limits on adaptation. **a**, Correlation between actual force field and internal model acquired after 50 targets by adaptive controllers with various width gaussians, versus the frequency of the nonlinear force field (m in equation (6)). **b**, A high spatial frequency field ($m = 4$) in which subjects and a simulated controller trained. **c**, The internal model learned by the controller with $\sigma = 0.12 \text{ m s}^{-1}$ gaussians after training for 50 movements. Note that for directions 22.5° and 157.5° , the controller predicts the field to be resistive, whereas actual forces are

assistive. **d**, Hand velocities parallel to target direction of three subjects trained in the same high-spatial-frequency field in the same target set as the simulated controller. Plotted are velocities in baseline null field movements (mean $\pm$ standard deviation) and during catch trials in targets 51 and 58 toward 22.5° and 157.5° . In catch trials, the hand is moving significantly faster, consistent with an internal model that expects a resistive field.

ϕ . The correlation between $\mathbf{b}(\phi)$ in the simulations and the subject data was strongest for $\sigma = 0.12 \text{ m s}^{-1}$.

We next used the model to predict behaviour beyond the data set with which the primitives were estimated. We noted that gaussian width influences how force estimation generalizes across both directions and speeds. Simulations predicted that when learning relied upon wide gaussians, reaching movements would not monotonically converge onto a straight line desired trajectory but would become S-shaped (Fig. 3a). Whereas the force field was linear in velocity, wide gaussians produced an approximation that overestimated forces at low speeds and underestimated forces at high speeds (Fig. 3b). Overestimation of the forces resulted in overcompensation of the field early in the movement; the magnitude of overcompensation depended on the gaussian width (Fig. 3c). With narrow gaussians, the simulated internal model did not overcompensate, but with wide gaussians movements became S-shaped. To test this prediction, we trained 24 subjects in target sets without catch trials. Movements of subjects were S-shaped (Fig. 3e), similar to movements made by simulations that learned with wide gaussians (Fig. 3a).

Simulations further predicted that the probability of catch trials influenced whether movements would become S-shaped (Fig. 3d). If catch trials occurred with 17% probability, then even with gaussians of $\sigma = 0.12 \text{ m s}^{-1}$ there should be sufficient unlearning caused by each catch trial such that hand trajectories would converge toward a straight line, without overcompensation (Fig. 3g). We tested this prediction by training subjects in target sets with 17% catch trial probability. As predicted, subjects did not show overcompensation (Fig. 3f).

Because approximation of a high-frequency signal with low-frequency bases generally results in poor representation, we next explored the limitations of a system that learns with wide gaussians. We simulated learning of nonlinear force fields:

$$\mathbf{f} = -13\sqrt{\dot{x}^2 + \dot{y}^2} \begin{bmatrix} -\sin(m\theta) \\ \cos(m\theta) \end{bmatrix} \quad (6)$$

$$\theta = \arctan(\dot{y}/\dot{x})$$

where $\dot{x}$ and $\dot{y}$ were components of hand velocity in Cartesian space. When $m = 1$, the field was the curl pattern¹³ learned by subjects above. As m increased, the field's spatial frequency increased. For various σ , we simulated movements to 50 targets and correlated the internal model learned by the simulation to the actual force field (Fig. 4a). As the field's spatial frequency increased, the accuracy of the internal model decreased for all basis function widths. However, the learning capability of wider bases collapsed at lower frequencies. This agrees with the recent finding that humans demonstrated a lesser ability to adapt in higher-spatial-frequency force fields²³.

To illustrate this deficiency, we trained an adaptive controller ($\sigma = 0.12 \text{ m s}^{-1}$) for 50 movements in a high-spatial-frequency field ($m = 4$, Fig. 4b). Because approximation is performed with wide bases, the internal model learned by the controller (Fig. 4c) cannot represent faithfully the rapidly changing forces. In particular, the simulation predicts that in movements toward 22.5° and 157.5° the internal model expects resistive forces where the force field is assistive. We tested for this prediction in three subjects by training each with the same random pattern of targets presented to the simulation, then presenting two catch trials toward 22.5° and 157.5°. Subjects behaved as though they were expecting a resistive field in those directions, as illustrated by their hand velocities in the catch trials (Fig. 4d).

Errors in learning dynamics of arm movements suggest that the brain composes motor commands with computational elements that are broadly tuned to arm velocity. When expressed in polar coordinates, the gaussians exhibit a preferred direction of movement, much like the cosine tuning curves typically associated with

cells in the motor system. However, our inferred bases have on average a half-width at half-height value of about 40°, significantly less than the 90° value required of cosines. Recent results²⁴ have demonstrated that tuning curves in monkey motor cortex have a median width of 56°, also much narrower than cosines. Motor cortical cells, however, have been reported to encode hand speed linearly¹⁵. Learning a linear force field with bases that have cosine directional tuning and linear speed tuning results in an internal model that does not produce the S-shaped movements observed in our subjects. Wide gaussians predict this unusual behaviour. They also explain why humans generalize sublinearly to fast movements after training in a linear force field at slow speeds⁸. The nonlinear encoding of speed inherent to gaussians resemble tuning properties of Purkinje cells that encode arm movements in the cerebellum²². Although several investigators have proposed a major role for the cerebellum in learning of internal models^{25–27}, our results suggest a link between patterns of generalization and firing properties of cells in this area. Since the output of the cerebellum partially affects the motor cortex²⁹, the finding that the preferred direction of motor cortical cells rotates during learning of force fields²⁸ may be a consequence of changing input from the cerebellum³⁰. □

Methods

Three groups of right-handed normal human subjects were trained to make movements while holding the handle of a lightweight robotic arm. All movements were toward a pseudorandomly chosen target, then back to a centre target. Targets appeared at 0°, 45°, ... 315°, at 10-cm displacement. Desired movement duration was $500 \pm 50 \text{ ms}$. Timing feedback was provided by changing the target colour. All subjects initially practised the task without any perturbing force (the null field). The first group of subjects ($n = 40$) trained in target sets of 192 movements during which the force field $\mathbf{B} = [0, 13; -13, 0] \text{ N s m}^{-1}$ was applied in 83% of the targets. The remaining 17% of targets (pseudorandomly selected) were catch trials during which the force field was turned off. In 13 of these subjects, EMG from anterior and posterior deltoid, biceps and triceps were measured with surface electrodes, amplified, filtered, r.m.s. (root mean square)-rectified, averaged from –50 to 100 ms into the movement, multiplied by unit vectors pointing toward movement direction, and summed across movement directions to produce a preferred direction for each muscle⁵. The second group of subjects ($n = 24$) trained in the same target sets as above, but with the force field always on (no catch trials). A final group of subjects ($n = 3$) trained for 58 movements in a higher-frequency force field (equation (6); $m = 4$), receiving catch trials only on targets 51 and 58 in directions 22.5° and 157.5°.

A simulated anthropomorphic controller¹³ made movements to the same sequence of targets experienced by subjects. The controller's internal model learned to map desired hand velocities into forces. Desired hand trajectories were minimum jerk, 0.5 s in duration, and 10 cm long. The internal model approximated the imposed force field (equation (1)) with gaussian bases (equation (5)); controllers were differentiated by σ . Gaussian centres spanned a range of desired velocities (–0.5 to 0.5 m s^{-1} in the x - and y -direction) spaced one σ apart. Sensitivity analyses suggested there were no significant changes in the results when the density of the bases was increased by an order of magnitude. Weights were initially randomized, then updated using equation (2) with $\eta = 0.0025$. To compare the output of gaussian basis functions with cosine tuning curves, the unweighted output of each gaussian was averaged during movement time. The collection of averaged outputs across movement directions formed a tuning curve for each gaussian primitive. Tuning curves were aligned to the preferred direction of each gaussian and averaged across all gaussians, from which the half-width at half-height value was calculated.

Received 17 May; accepted 5 July 2000.

1. Mussa-Ivaldi, F. A., Giszter, S. F. & Bizzi, E. Linear combinations of primitives in vertebrate motor control. *Proc. Natl Acad. Sci. USA* **91**, 7534–7538 (1994).
2. Schaal, S. & Atkeson, C. G. Constructive incremental learning from only local information. *Neural Comput.* **10**, 2047–1084 (1998).
3. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
4. Ghez, C., Krakauer, J. W., Sainburg, R. L. & Ghilardi, M. F. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 501–514 (MIT Press, Cambridge, Massachusetts, 2000).
5. Thoroughman, K. A. & Shadmehr, R. Electromyographic correlates of learning internal models of reaching movements. *J. Neurosci.* **19**, 8573–8588 (1999).
6. Gandolfo, F., Mussa-Ivaldi, F. A. & Bizzi, E. Motor learning by field approximation. *Proc. Natl Acad. Sci. USA* **93**, 3843–3846 (1996).
7. Condit, M. A., Gandolfo, F. & Mussa-Ivaldi, F. A. The motor system does not learn the dynamics of the arm by rote memorization of past experience. *J. Neurophysiol.* **78**, 554–560 (1997).
8. Goodbody, S. J. & Wolpert, D. M. Temporal and amplitude generalization in motor learning. *J. Neurophysiol.* **79**, 1825–1838 (1998).
9. Condit, M. A. & Mussa-Ivaldi, F. A. Central representation of time during motor learning. *Proc. Natl Acad. Sci. USA* **96**, 11625–11630 (1999).
10. Kawato, M. Adaptation and learning in control of voluntary movement by the central nervous system. *Adv. Robotics* **3**, 229–249 (1989).

11. Sanner, R. M. & Kosha, M. A mathematical model of the adaptive control of human arm motions. *Biol. Cybern.* **80**, 369–382 (1999).
12. Bhushan, N. & Shadmehr, R. Computational architecture of human adaptive control during learning of reaching movements in force fields. *Biol. Cybern.* **81**, 39–60 (1999).
13. Shadmehr, R. & Brashers-Krug, T. Functional stages in the formation of human long-term motor memory. *J. Neurosci.* **17**, 409–419 (1997).
14. Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. Neural population coding of movement direction. *Science* **233**, 1416–1419 (1986).
15. Moran, D. W. & Schwartz, A. B. Motor cortical representation of speed and direction during reaching. *J. Neurophys.* **82**, 2676–2692 (1999).
16. Conrad, B., Matsunami, K., Meyer-Lohmann, J., Wiesendanger, M. & Brooks, V. B. Cortical load compensation during voluntary elbow movements. *Brain Res.* **71**, 507–514 (1974).
17. Goodkin, H. P., Keating, J. G., Martin, T. A. & Thach, W. T. Preserved simple and impaired compound movement after infarction in the territory of the superior cerebellar artery. *Can. J. Neurol. Sci.* **20**, S93–S104 (1993).
18. Bastian, A. J., Martin, T. A., Keating, J. G. & Thach, W. T. Cerebellar ataxia: abnormal control of interaction torques across multiple joints. *J. Neurophysiol.* **76**, 492–509 (1996).
19. Lang, C. E. & Bastian, A. J. Cerebellar subjects show impaired adaptation of anticipatory EMG during catching. *J. Neurophysiol.* **82**, 2108–2119 (1999).
20. Nezafat, R., Shadmehr, R. & Holcomb, H. H. Neural correlates of motor memory retention over a 4 week period as quantified through PET. *Soc. Neurosci. Abs.* **25**, 2178 (1999).
21. Imamizu, H. *et al.* Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature* **403**, 192–195 (2000).
22. Coltz, J. D., Johnson, M. T. V. & Ebner, T. J. Cerebellar Purkinje cell simple spike discharge encodes movement velocity in primates during visuomotor arm tracking. *J. Neurosci.* **19**, 1782–1803 (1999).
23. Matsouka, Y. Models of generalization in motor control. PhD thesis, MIT (1998).
24. Amirikian, B. & Georgopoulos, A. P. Directional tuning profiles of motor cortical cells. *Neurosci. Res.* **36**, 73–79 (2000).
25. Houk, J. C. & Wise, S. P. Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb. Cortex* **5**, 95–110 (1995).
26. Wolpert, D. M. & Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Networks* **11**, 1317–1329 (1998).
27. Spoelstra, J., Schweighofer, N. & Arbib, M. A. Cerebellar learning of accurate predictive control for faster-reaching movements. *Biol. Cybern.* **82**, 321–333 (2000).
28. Gandolfo, F., Li, C. S. R., Benda, B. J., Schioppa, C. P. & Bizzi, E. Cortical correlates of learning in monkeys adapting to a new dynamical environment. *Proc. Natl Acad. Sci. USA* **97**, 2259–2263 (2000).
29. Holdefer, R. N., Miller, L. E. & Houk, J. C. Functional connectivity between cerebellum and primary motor cortex in the awake monkey. *J. Neurophysiol.* **84**, 585–590 (2000).
30. Martin, J. H., Cooper, S. E., Hacking, A. & Ghez, C. Differential effects of deep cerebellar nuclei inactivation on reaching and adaptive control. *J. Neurophysiol.* **84**, 1886–1899 (2000).

Acknowledgements

We thank M. A. Smith, O. Donchin, and R. Nezafat. The work was supported by grants from the Office of Naval Research and the National Institutes of Health (to R.S.), and a research traineeship from the National Science Foundation (K.A.T.).

Correspondence should be addressed to R.S. (e-mail: reza@bme.jhu.edu).

Netrin-1-mediated axon outgrowth and cAMP production requires interaction with adenosine A2b receptor

Véronique Corset*, Kim Tuyen Nguyen-Ba-Charvet†, Christelle Forcet*, Emmanuel Moysse*, Alain Chédotal† & Patrick Mehlen*

* Laboratoire Apoptose et Différenciation, label 'La Ligue' – Centre de Génétique Moléculaire et Cellulaire, CNRS UMR 5534, Université Lyon 1, 69622 Villeurbanne, France

† INSERM U106, Hôpital de la salpêtrière, 75651 Paris cedex 13, France

The netrins, a family of laminin-related secreted proteins, are critical in controlling axon elongation and pathfinding^{1–4}. The DCC (for deleted in colorectal cancer) protein was proposed as a receptor for netrin-1 in the light of many observations including the inhibition of netrin-1-mediated axon outgrowth and attraction in the presence of an anti-DCC antiserum^{5–7}, the similitude of nervous system defects in DCC and netrin-1 knockout mice^{4,8} and the results of receptor swapping experiments⁹. Previous studies

have failed to show a direct interaction of DCC with netrin-1 (ref. 10), suggesting the possibility of an additional receptor or co-receptor. Here we show that DCC interacts with the membrane-associated adenosine A2b receptor, a G-protein-coupled receptor that induces cAMP accumulation on binding adenosine¹¹. We show that A2b is actually a netrin-1 receptor and induces cAMP accumulation on binding netrin-1. Finally, we show that netrin-1-dependent outgrowth of dorsal spinal cord axons directly involves A2b. Together our results indicate that the growth-promoting function of netrin-1 may require a receptor complex containing DCC and A2b.

To identify direct partners of DCC, we carried out a two-hybrid screen of a human brain complementary DNA library using the intracellular domain of DCC as bait. The screen revealed the putative binding of different proteins to the intracellular domain of DCC. Some of these partners, like caspase-3, have been described elsewhere^{12,13}. We also observed putative binding of actin and dynactin to DCC (data not shown); this could be important because actin and dynactin are cytoskeleton proteins that are involved in neurite elongation and guidance. But here we focused our attention on the relevance of the DCC binding to a 23-residue fragment corresponding to the last 23 amino acids (310–333) of the intracellular domain of the adenosine A2b receptor (Fig. 1a). A2b is a member of the family of the adenosine-specific receptors¹¹, a family which was originally divided into four subtypes—the A1, A2a, A2b and A3 receptors—on the basis of their ability to inhibit (through A1 or A3) or stimulate (through A2a or A2b) adenylate cyclase activity in response to adenosine binding¹⁴. The A2b receptor couples to either G_q or G_s, leading to a phospholipase C-dependent pathway¹⁵ or an accumulation of cAMP¹⁶, respectively. The low affinity of the A2b receptor for adenosine as compared with A2a questioned the role of this receptor *in vivo*.

We confirmed the DCC/A2b interaction identified in the two-hybrid screen by co-immunoprecipitation (Fig. 1b). In 293T cells co-transfected with A2b and DCC, a DCC/A2b interaction was barely detectable in the absence of netrin-1, whereas there was a very strong binding when netrin-1 was present (Fig. 1b). Hence, DCC

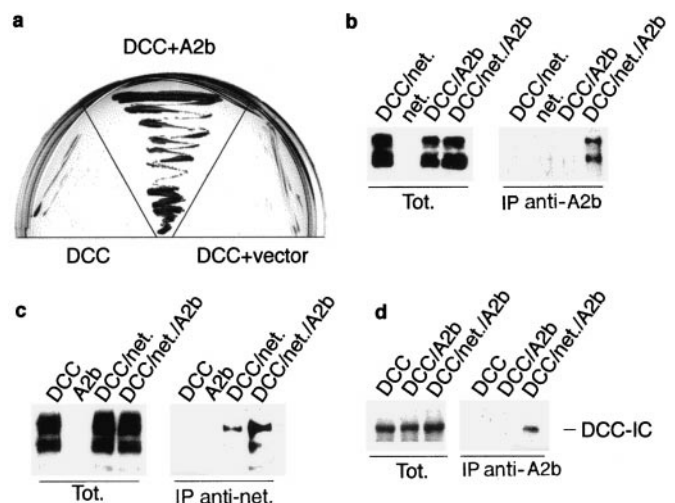


Figure 1 DCC interacts with the adenosine A2b receptor. **a**, Two-hybrid screen of DCC-IC revealed interaction with A2b. Transformed Y190 yeast growth in the selection medium is shown. **b–d**, Co-immunoprecipitations were performed on 293T cells co-transfected with combinations of the A2b encoding plasmid PRC/CMV-A2b-Flag (A2b), the full-length DCC encoding plasmid pDCC-CMV-S (DCC) and the netrin-1 encoding plasmid pGNET1myc (net). The pull-down assays were done with anti-Flag (**b**, **d**) or anti-c-Myc (**c**) antibody. In **d**, DCC-IC fused to a myristoylation signal¹² was expressed instead of DCC full-length. 'tot.' and 'IP' indicate DCC immunoblots before and after the pull-down assays, respectively.

requires the presence of netrin-1 to interact with the A2b receptor. A similar netrin-1-gated effect is necessary for the binding of DCC to another class of netrin receptors, the vertebrate UNC-5 homologues⁹. We next analysed whether the A2b presence modulated DCC/netrin-1 interaction. In 293T cells, we observed a detectable pull-down of DCC with netrin-1, suggesting that DCC/netrin-1 interaction takes place (a nonspecific interaction cannot be excluded, however, as netrins are heparin-binding proteins that can associate nonspecifically with cell membranes⁵). The co-immunoprecipitation of DCC by netrin-1 pull-down was fivefold greater when A2b was co-expressed (Fig. 1c); and when we carried out a co-immunoprecipitation study using only the intracellular domain of DCC (DCC-IC), the presence of netrin-1 also resulted in the co-immunoprecipitation of DCC-IC by A2b (Fig. 1d), suggesting that DCC mainly interacted indirectly with netrin-1 through the A2b receptor.

To analyse further the netrin-1/A2b interaction, we carried out co-immunoprecipitations to assess DCC-independent netrin-1 binding to the A2b receptor. Figure 2a shows that netrin-1 did interact with A2b in transfected 293T cells; notably, the A2b/netrin-1 interaction was not modulated by DCC expression (Fig. 2a). Similarly, immunocytochemical analysis also suggested that netrin-1 binds to the A2b receptor (Fig. 2b). Specific membrane-bound netrin-1 was barely detectable in 293T cells expressing DCC unless

heparin, which minimizes background binding, was used (ref. 5; data not shown), whereas specific signal was observed in cells expressing A2b receptor even in the absence of heparin (Fig. 2b). We determined the equilibrium binding constant using ¹²⁵I-labelled netrin-1 (Fig. 2c); the K_d (11 nM) represented an adequate value considering the physiological concentration of netrin-1 required for axon outgrowth (low nanomolar range). Biochemical characterization of the netrin-1 interaction with radioligand-complexed A2b revealed that netrin-1 does not interact with the adenosine-binding site of A2b, but potentiates radioligand-specific binding (Fig. 2d). This result suggests that the A2b receptor displays a specific binding site for netrin-1 that is distinct from the classical adenosine-binding site, like GABA receptors for benzodiazepines^{17,18}.

The growth cone attraction mediated by netrin-1 is modulated by intracellular cAMP concentration¹⁹, and netrin-1 induces an increase of cAMP in responding growth cones²⁰. We therefore thought that the A2b receptor was a candidate for modulating cAMP concentration in response to netrin-1. 293T cells expressing A2b showed a marked increase in cAMP levels in the presence of netrin-1 or the adenosine agonist NECA (5-*N*-ethylcarboxamido adenosine)²¹ (Fig. 3a). Moreover, in agreement with the binding data (Fig. 2d), expression of netrin-1 in the presence of NECA further increased cAMP production. The addition of the specific A2b antagonist enprofylline (3-*n*-propylxanthine)²² completely reversed the increase of cAMP level induced by netrin-1 (Fig. 3a). A similar block was observed when another adenosine receptor antagonist DPCPX (8-cyclopentyl-1,3-dipropylxanthine, 1 μ M) was used (data not shown). Moreover, netrin-1-induced cAMP production through A2b was independent of extracellular adenosine or ATP that netrin-1 could have released independently (Fig. 3b). Hence, netrin-1 is a direct stimulator of cAMP production through A2b. Notably, 0.5–5 nM netrin-1, concentrations sufficient to stimulate axon outgrowth *in vivo*, induced cAMP production (Fig. 3b), suggesting that physiological concentrations of netrin-1 stimulate the A2b-induced increase in cAMP. cAMP antagonists can convert *Xenopus* growth cone attraction to repulsion^{19,23}. Thus, netrin-1-induced production of cAMP through the A2b

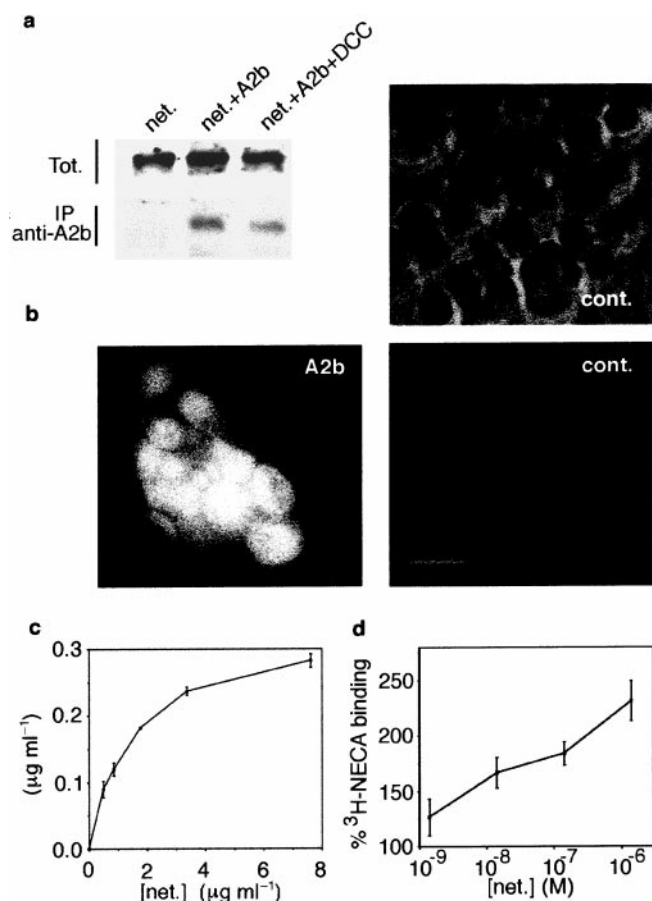


Figure 2 A2b is a netrin-1 receptor. **a**, Netrin-1/A2b co-immunoprecipitations were performed on 293T transfected cells (see legend to Fig. 1b–d) using anti-Flag M2 and anti-c-Myc antibodies for the pull-down and the immunoblot, respectively. **b**, Netrin-1 immunostaining on mock-transfected cells (cont.) versus A2b transfected cells (A2b). A phase contrast image of the mock cells (cont.) is also shown. Scale bar, 10 μ m. **c**, Equilibrium binding of netrin-1 to A2b expressing cells. The quantification of specifically bound ¹²⁵I-labelled netrin-1 versus free ¹²⁵I-labelled netrin-1 is shown. The K_d was deduced from the Scatchard plot. **d**, Netrin-1 potentiates NECA binding. ³H-NECA binding to A2b-containing membranes after addition of various netrin-1 concentrations is shown.

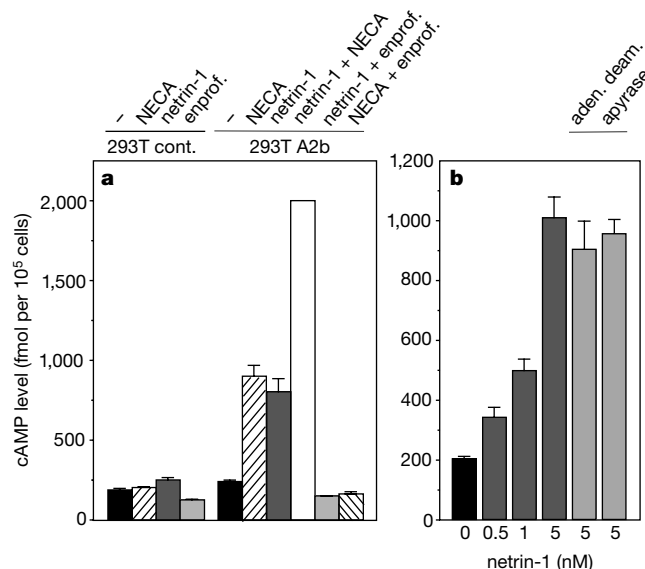


Figure 3 Netrin-1 mediates cAMP production through the A2b receptor. **a**, 293T cells were transfected with the control plasmid, the A2b encoding plasmid PRC/CMV-A2b (A2b) or the netrin-1 encoding plasmid pGNET1myc (net.) as indicated. NECA (100 μ M) and enprofylline (300 μ M) were added 24 h after transfection; and 24 h later the cAMP level was monitored (see Methods). **b**, Different concentrations of purified netrin-1 were added to A2b transfected cells for 3 h in presence or in absence of apyrase (1 U ml⁻¹) or adenosine deaminase (1 U ml⁻¹), and cAMP levels were measured.

receptor might represent a key signalling event to allow growth cone attraction.

Because netrin-1 promoted axon outgrowth for several classes of neurons, we investigated whether A2b was implicated in this process, using netrin-1-mediated outgrowth of commissural axons in the embryonic dorsal spinal cord^{5,24}. We grew dorsal spinal cord explants from E11 rat embryos for 36–40 h in collagen gel, adjacent to mock-transfected COS cells, or netrin-1-expressing COS cells in the presence or absence of the selective antagonist enprofylline²². The addition of enprofylline in the 100–600 μ M concentration range was sufficient to block significantly netrin-1-induced growth cone extension ($P < 0.0001$; Fig. 4). The use of 1 μ M DPCPX also strongly reduced netrin-1-dependent axonal outgrowth (data not shown). In all enprofylline-treated explants, at least 90% of the few axons that were still able to grow were facing netrin-1-expressing cells (Fig. 4c), suggesting that the attractive effect of netrin-1 might not be totally affected. In contrast, netrin-1-independent axonal outgrowth, which can be observed when dorsal spinal cord explants are cultured more than 40 h (refs 2, 24), was not affected by 300 μ M enprofylline (Fig. 4d). Moreover, axons growing from the explants in the presence of netrin-1 could be stained using an anti-A2b antibody (Fig. 4g, h). These results, combined with the *in vitro* analysis, which showed that netrin-1 interacted with A2b and the A2b-dependent netrin-1 induced cAMP production, indicate that netrin-1 function requires the A2b receptor.

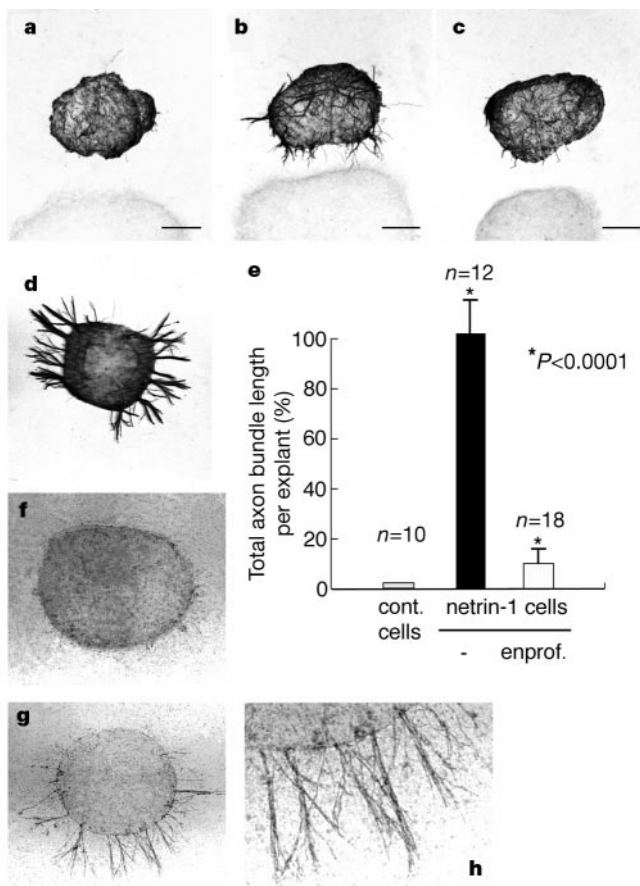


Figure 4 A2b is required for netrin-1-dependent outgrowth of dorsal spinal cord axons. **a–d**, E11 dorsal spinal cord explants were cultured for 40 h (**a–c**) or 72 h (**d**) in collagen gel next to mock-transfected COS cells (**a**), or COS cells transfected with pGNET1 myc, in the absence (**b**) or presence (**c**, **d**) of 300 μ M of enprofylline. **e**, Quantification of the results shown in **a–c**. *n* is the total number of explants that were quantified from four distinct experiments. Values shown are means $\pm$ s.e.m. Scale bars, 170 μ m. Asterisk, $P < 0.0001$. **f–h**, E11 explants cultured as in **b** were immunostained with **g**, magnification $\times 2.3$ shown in **h**) or without (**f**) A2b antibody.

Together, these results suggest that A2b is a receptor for netrin-1. They also reveal the first connection between the extracellular cue netrin-1 and an intracellular cAMP accumulation. Whether other axon guidance cues, also suggested²⁵ to activate a cAMP-dependent signalling pathway, interact with adenosine-receptors remains to be examined.

Although our data clearly demonstrate an interaction between netrin-1 and A2b, and the role of A2b in netrin-1-mediated axonal outgrowth, a direct interaction of DCC with netrin-1 cannot be excluded and might be important to help specify the response to netrin-1, in particular its function in axon guidance. DCC may then work in two different ways: first, as a co-receptor that, when engaged by the netrin-1/A2b complex, promotes axon outgrowth and/or attraction; and second, as a receptor for netrin-1 or another known (for example heparin²⁶) or unknown ligand leading to a multi-component complex with several receptors and ligands. In the latter case, the A2b/netrin-1 interaction would drive the cAMP-dependent pathway.

As netrin-1 binds to the A2b receptor on a site distinct from the adenosine site but also induces, through A2b stimulation, a massive cAMP production, adenosine may not be the only ligand for the adenosine receptor family. It might also explain why the A2b receptor increased cAMP in brain slices but apparently not in brain membranes²⁷. The implication of the A2b receptor in mediating netrin-1-dependent axon outgrowth indicates that adenosine receptors may have a function in the nervous system, far away from their known activity in neurotransmission. □

Methods

Cells, transfection procedures and immunoblotting

We carried out transient transfections of human embryonic kidney 293T cells as described¹², according to a modified calcium phosphate procedure. We analysed DCC and netrin-1 expression in 293T cells by immunoblot analysis as described¹² using anti-DCC (extracellular domain epitope from Oncogene Research Products and intracellular domain epitope from Santa Cruz) or anti-c-Myc 9E10 (Boehringer Mannheim) antibody.

Site-directed mutagenesis and plasmid constructs

Prc/CMV-A2b encodes the full-length A2b under the control of CMV promoter and was provided by P. Schofield. The pGNET1 myc encoding chick netrin has been described¹². We introduced the Flag epitope in the amino terminus of A2b using the quickchange strategy (Stratagene) using Prc/CMV-A2b as template and the following primers: 5'-GCTGGCC CGGCATGGATTACAAGGACGATGACAAGCTGCTGGAGACACAG-3' and 5'-CTGT GTCTCCAGCAGCTTGTTCATCGTCTCTGTAATCCATGGCCGGGCCAGC-3'.

Two-hybrid analysis

We used the two-hybrid system matchmaker II (Clontech) according to the manufacturer's instructions. As bait, we inserted the coding sequence of DCC-IC (1,121–1,447) from pSVK3-HA-DCC-IC (ref. 12) into pAS21 through *NdeI*–*SalI* insertion. Y190 yeast transformed with the GAL4 DNA binding domain/DCC-IC fusion were further transformed with plasmids containing the GAL4 transcriptional activation domain AD fused to a human brain cDNA library (Clontech). We grew cells in the absence of leucine, tryptophan and histidine and in the presence of 55 mM 3-amino-1,2,4-triazole.

Immunoprecipitation

We carried out co-immunoprecipitations on co-transfected 293T lysed in 50 mM HEPES pH 7.6, 125 mM NaCl, 5 mM EDTA and 0.1% NP-40 in the presence of protease inhibitor, using either an anti-DCC, Flag M2 or c-Myc 9E10 antibody and protein-A Sepharose (Sigma) to pull down DCC, A2b or netrin-1, respectively. To obtain similar expression of the protein DCC in the different samples, we co-transfected half of the plasmid pDCC–CMV-S with the netrin-1-expressing vector as the presence of netrin-1 leads to increased stability of DCC¹². We monitored DCC interaction with A2b or netrin-1 by immunoblot using anti-DCC antibody. Netrin-1 interaction with A2b was detected by immunoblot using the anti-c-Myc 9E10 antibody.

Binding experiments

Netrin-1 binding on transfected 293T cells was done as described⁸. Briefly, 293T cells transfected with mock or A2b-encoding plasmid were incubated in absence of heparin with 2 μ g ml⁻¹ purified netrin-1 for 1 h. After three PBS washes, cells were fixed with paraformaldehyde and incubated with anti-c-Myc antibody. FITC-labelled secondary antibody was then used to detect netrin-1. Bound netrin-1 was examined and photographed with a Zeiss Axioskop photomicroscope.

Equilibrium binding experiments

We carried out equilibrium binding experiments as described⁵. Briefly, ¹²⁵I-labelled netrin-1 was prepared using the iodo-beads procedure (Pierce) and subsequent purification using the D-Salt Dextran Desalting column (Pierce) allowing 84% recovery of reacted netrin-1. Mock or A2b-transfected 292T cells (2×10^5) were then incubated in PBS containing $2 \mu\text{g ml}^{-1}$ heparin and 1 mg ml^{-1} BSA with serial dilutions of ¹²⁵I-labelled netrin-1 for 3 h at 4 °C. We determined specifically bound netrin-1, calculated from the difference between A2b-expressing cells and mock cells, versus unbound netrin-1 by measuring the radioactivity of the cell pellets, after three PBS washes, and the respective incubation supernatants.

A2b radioligand binding

Membranes of cells transfected with A2b receptor were semi-purified by differential centrifugation into a P2 pellet, and resuspended in 50 mM Tris-HCl pH 7.4 (ref. 28). Similar membrane preparation was applied to freshly dissected rat striatum as positive control. Membrane suspensions were pre-incubated 30 min at 37 °C with 2 U ml^{-1} adenosine deaminase (Roche Diagnostic) to remove endogenous ligand. Radioligand binding assays were conducted at 21 °C for 3 h in 100 μl buffer containing 50 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 1 U ml^{-1} adenosine deaminase, 40 μg membrane proteins and 50 nM ³H-NECA (26 Ci mmol⁻¹, Amersham). Nonspecific binding, in the presence of 10 μM theophyllin or NECA, equalled 10% of total radioligand binding. We assayed ³H-NECA-binding in concentrations of purified netrin-1 ranging from 0.1 nM to 2 μM . Incubation was stopped by adding 1 ml of cold buffer, and mixtures were filtered on pre-soaked GF/C Whatman filters for membrane-bound radioactivity measurement in a liquid scintillation counter (Beckmann).

Determination of cAMP level

We measured the concentration of cAMP by using the biotrac cAMP enzyme immunoassay system (Amersham Pharmacia) according to the manufacturer's instructions. We calculated the cAMP level of co-transfected 293T cells using a standard curve ranging from 10 to 10^4 fmol of cAMP.

Dorsal spinal cord cultures

Dorsal spinal cord explants from E11 rat embryos were cultured as described²⁴, except that the culture medium for explants contained 100 μM NaOH (as the A2b-specific antagonist enprofylline was dissolved in NaOH). Axons were stained with an anti- β -tubulin antibody (Babco) or anti-rat A2b antibody (Santa Cruz). Quantification of axonal length was done as described²⁵. Briefly, the total length of axon bundles was measured for each explant and normalized to the values obtained from explants cultured with netrin-1 cells in presence of the enprofylline diluent.

Received 26 June; accepted 25 July 2000.

1. Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
2. Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425–435 (1994).
3. Colamarino, S. A. & Tessier-Lavigne, M. The axonal chemoattractant netrin-1 is also a chemorepellent for trochlear motor axons. *Cell* **81**, 621–629 (1995).
4. Serafini, T. et al. Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
5. Keino-Masu, K. et al. Deleted in colorectal cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175–185 (1996).
6. Deiner, M. S. et al. Netrin-1 and DCC mediate axon guidance locally at the optic disc: loss of function leads to optic nerve hypoplasia. *Neuron* **19**, 575–589 (1997).
7. de la Torre, J. R. et al. Turning of retinal growth cones in a netrin-1 gradient mediated by the netrin receptor DCC. *Neuron* **19**, 1211–1224 (1997).
8. Fazeli, A. et al. Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene. *Nature* **386**, 796–804 (1997).
9. Hong, K. et al. A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell* **97**, 927–941 (1999).
10. Meyerhardt, J. A. et al. Netrin-1: interaction with deleted in colorectal cancer (DCC) and alterations in brain tumors and neuroblastomas. *Cell Growth Differ.* **10**, 35–42 (1999).
11. Feoktistov, I. & Biaggioni, I. Adenosine A2b receptors. *Pharmacol. Rev.* **49**, 381–402 (1997).
12. Mehlen, P. et al. The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. *Nature* **395**, 801–804 (1998).
13. Forcet, C. et al. The dependence receptor DCC-mediated mitochondria independent caspase-9 dependent caspase activation. *Proc. Natl Acad. Sci. USA*, submitted.
14. Ralevic, V. & Burnstock, G. Receptors for purines and pyrimidines. *Pharmacol. Rev.* **50**, 413–492 (1998).
15. Feoktistov, I. & Biaggioni, I. Adenosine A2b receptors evoke interleukin-8 secretion in human mast cells. An enprofylline-sensitive mechanism with implications for asthma. *J. Clin. Invest.* **96**, 1979–1986 (1995).
16. Yakel, J. L., Warren, R. A., Reppert, S. M. & North, R. A. Functional expression of adenosine A2b receptor in *Xenopus* oocytes. *Mol. Pharmacol.* **43**, 277–280 (1993).
17. Squires, R. F. et al. Honokiol and magnolol increase the number of [³H]muscimol binding sites threefold in rat forebrain membranes in vitro using a filtration assay, by allosterically increasing the affinities of low-affinity sites. *Neurochem. Res.* **24**, 1593–1602 (1999).
18. Boileau, A. J., Evers, A. R., Davis, A. F. & Czajkowski, C. Mapping the agonist binding site of the GABAA receptor: evidence for a beta-strand. *J. Neurosci.* **19**, 4847–4854 (1999).
19. Ming, G. L. et al. cAMP-dependent growth cone guidance by netrin-1. *Neuron* **19**, 1225–1235 (1997).
20. Hopker, V. H., Shewan, D., Tessier-Lavigne, M., Poo, M. & Holt, C. Growth-cone attraction to netrin-1 is converted to repulsion by laminin-1. *Nature* **401**, 69–73 (1999).

21. Feoktistov, I. & Biaggioni, I. Characterization of adenosine receptors in human erythroleukemia cells and platelets: further evidence for heterogeneity of adenosine A2 receptor subtypes. *Mol. Pharmacol.* **43**, 909–914 (1993).
22. Fredholm, B. B. & Persson, C. G. Xanthine derivatives as adenosine receptor antagonists. *Eur. J. Pharmacol.* **81**, 673–676 (1982).
23. Song, H. et al. Conversion of neuronal growth cone responses from repulsion to attraction by cyclic nucleotides. *Science* **281**, 1515–1518 (1998).
24. Serafini, T. et al. The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. *Cell* **78**, 409–424 (1994).
25. Song, H. & Poo, M. Signal transduction underlying growth cone guidance by diffusible factor. *Curr. Opin. Neurobiol.* **9**, 355–363 (1999).
26. Bennett, K. L. et al. Deleted in colorectal carcinoma (DCC) binds heparin via its fifth fibronectin type III domain. *J. Biol. Chem.* **272**, 26940–26946 (1997).
27. Daly, J. W., Butts-Lamb, P. & Padgett, W. Subclasses of adenosine receptors in the central nervous system: interaction with caffeine and related methylxanthines. *Cell. Mol. Neurobiol.* **3**, 69–80 (1983).
28. Linden, J., Thai, T., Figler, H., Jin, X. & Robeva, A. S. Characterization of human A(2B) adenosine receptors: radioligand binding, western blotting, and coupling to G(q) in human embryonic kidney 293 cells and HMC-1 mast cells. *Mol. Pharmacol.* **56**, 705–713 (1999).
29. de Castro, F., Hu, L., Drabkin, H., Sotelo, C. & Chedotal, A. Chemoattraction and chemorepulsion of olfactory bulb axons by different secreted semaphorins. *J. Neurosci.* **19**, 4428–4436 (1999).

Acknowledgements

We thank L. Granger for technical assistance, P. Schofield for the human A2b expressing construct, and M. Tessier-Lavigne for the purified netrin-1. This work was supported by the CNRS, the Ligue Contre le Cancer, the FRM, the ARC (to P.M. and to A.C.), the 'emmergence' program (to P.M.), and APEX INSERM (to A.C.). K.T.N.-B.-C is a recipient of a postdoctoral fellowship from IRP.

Correspondence and requests for materials should be addressed to P.M. (e-mail: mehlen@univ-lyon1.fr).

Somatic support cells restrict germline stem cell self-renewal and promote differentiation

Amy A. Kiger*, Helen White-Cooper† & Margaret T. Fuller*‡

* Department of Developmental Biology and ‡ Department of Genetics, Stanford University School of Medicine, Stanford, California 94305-5329, USA
† Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

Stem cells maintain populations of highly differentiated, short-lived cell-types, including blood, skin and sperm, throughout adult life¹. Understanding the mechanisms that regulate stem cell behaviour is crucial for realizing their potential in regenerative medicine². A fundamental characteristic of stem cells is their capacity for asymmetric division: daughter cells either retain stem cell identity or initiate differentiation. However, stem cells are also capable of symmetric division where both daughters remain stem cells^{3–6}, indicating that mechanisms must exist to balance self-renewal capacity with differentiation. Here we present evidence that support cells surrounding the stem cells restrict self-renewal and control stem cell number by ensuring asymmetric division. Loss of function of the *Drosophila* Epidermal growth factor receptor in somatic cells disrupted the balance of self-renewal versus differentiation in the male germline, increasing the number of germline stem cells. We propose that activation of this receptor specifies normal behaviour of somatic support cells; in turn, the somatic cells play a guardian role, providing information that prevents self-renewal of stem cell identity by the germ cell they enclose.

The identity of *Drosophila* male germline stem cells and their physical relationship with surrounding somatic cells are known, providing an excellent model for genetic analysis of stem cell

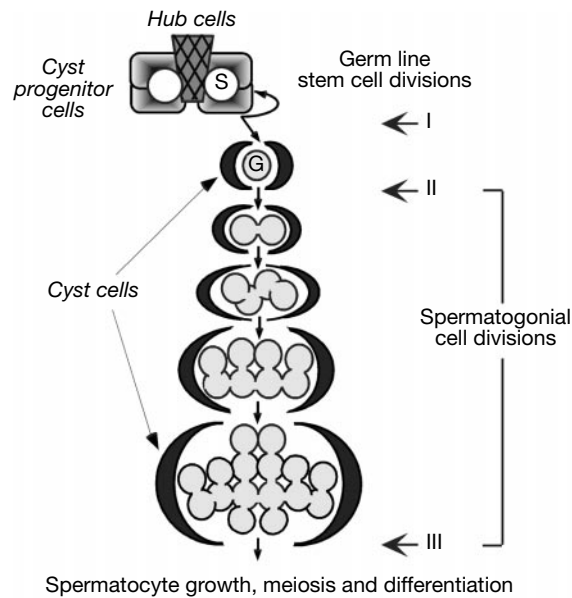


Figure 1 Three critical decision points in *Drosophila* early male germ cell differentiation. Germline stem cells (S) and somatic cyst progenitor stem cells (grey) surround somatic hub cells (hatched grey) at the testis apical tip. Arrow I, asymmetric germline stem cell division produces a new stem cell and a gonialblast (G), which becomes enclosed by two

somatic cyst cells (black). Arrow II, gonialblast initiates four rounds of mitotic amplification divisions with incomplete cytokinesis. Arrow III, cysts of sixteen interconnected spermatogonia exit mitosis and initiate spermatocyte growth, meiosis and terminal differentiation.

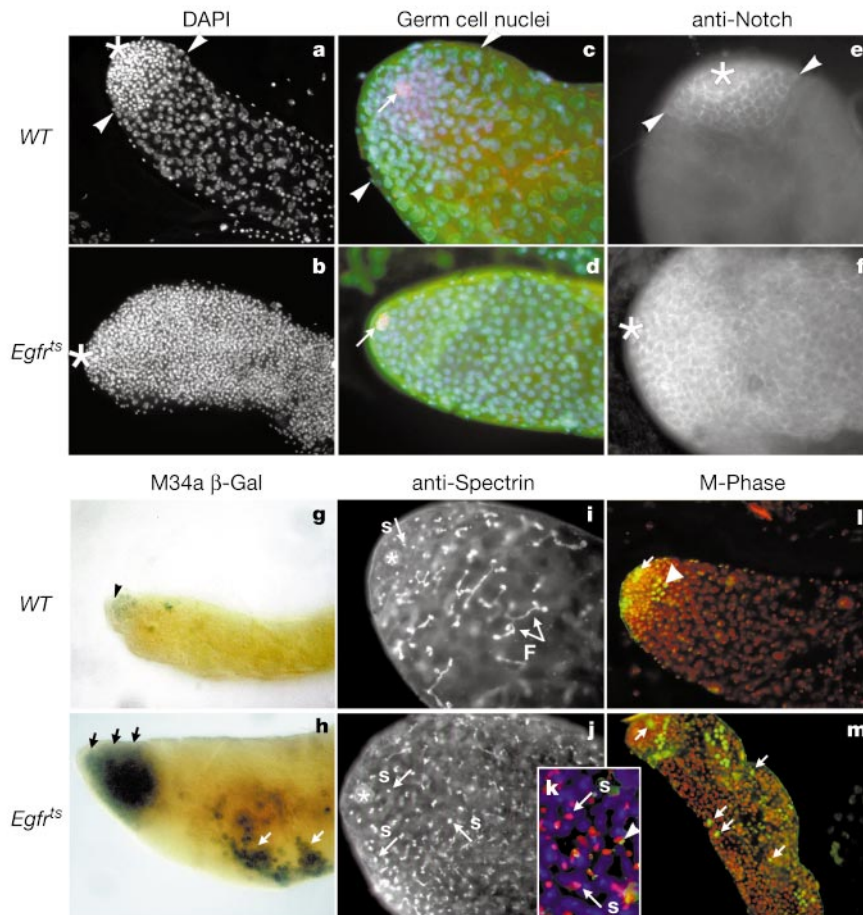


Figure 2 *Egfr^{ts}* restricts proliferation of early germ cells. Cytology of testes from wild-type (WT; **a, c, e, g, i, l**) and *Egfr^{ts}* (**b, d, f, h, j, k, m**) *Drosophila*. Asterisks, apical tips of testes. **a, b**, DAPI-staining. Arrowheads, spermatogonia/spermatocyte transition. **c, d**, Nuclear size increases with differentiation in control (**c**) but not *Egfr^{ts}* (**d**) testes. Green, anti-Vasa, germline nuclei; red and arrow, anti-FasIII, hub; blue, DAPI. **e, f**, Anti-Notch staining. **g, h**, Staining for M34a. Arrowhead, wild-type stem cells and gonialblasts;

black arrow, marked cells at *Egfr^{ts}* tip; white arrow, marked cells displaced from *Egfr^{ts}* tip. **i-k**, Anti- α -Spectrin staining. Arrow, spectrosomes; forked arrows, fusomes; green and arrowhead (**k**), few ring canals²⁸ in *Egfr^{ts}* displaced cells with spectrosomes. **l, m**, Staining with anti-phosphorylated-Histone-3 (green) and DAPI (red). Arrowhead, synchronous wild-type spermatogonial division; arrow, single dividing cell.

regulatory mechanisms. Male germline stem cells lie at the testis apical tip, in intimate contact with somatic hub cells and enveloped by somatic stem cells termed cyst progenitors (Fig. 1)⁷. Germline stem cell divisions normally have asymmetrical outcomes: the daughter adjacent to the hub retains stem cell identity and self-renewal capacity, while the daughter displaced from the hub becomes a gonialblast and initiates differentiation (Fig. 1, arrow I)^{7,8}. Cyst progenitors also divide asymmetrically to self-renew and produce a cyst cell, two of which enclose each gonialblast and its progeny⁷⁻⁹. The gonialblast initiates synchronous mitoses with incomplete cytokinesis, producing interconnected spermatogonia (Fig. 1, arrow II). After these four amplifying divisions, the resulting sixteen spermatogonia cease mitosis and enter meiosis as spermatocytes (Fig. 1, arrow III). Progression of germline differentiation occurs in a spatial gradient, with small, mitotic cells at the apex and larger, differentiating spermatocytes more distal (Fig. 2a, c).

The asymmetric outcome of stem cell divisions could result from differential inheritance of intrinsic determinants, exposure to extrinsic signals, or both. We found that *Egfr*⁺ function in somatic cells is required to restrict self-renewal of *Drosophila* male germline stem cells and allow differentiation, indicating a role for extrinsic signals. Adult males with a temperature-sensitive mutation in *Egfr* (*Egfr*^{ts} mutants)¹⁰ were raised at permissive temperature then shifted to restrictive temperature. Although all stages of spermatogenesis were initially present, by one week at non-permissive temperature, testes were filled with massive numbers of early germ cells with small, Vasa-positive nuclei (Fig. 2b, d). Notch protein was normally detected in stem cells, gonialblasts and spermatogonia but not in spermatocytes (Fig. 2e). In *Egfr*^{ts} testes, the number of Notch-positive cells was greatly increased (Fig. 2f). The early germ cells appeared to proliferate at the expense of differentiation, as spermatocytes and spermatids were absent or markedly reduced in number.

Many early germ cells in *Egfr*^{ts} mutant testes displayed expression markers, subcellular structures and cell division behaviour characteristic of stem cells or gonialblasts. The *M34a*⁹ insert drives β -Galactosidase expression in germline stem cells and gonialblasts at the testis tip (Fig. 2g). In all *Egfr*^{ts} mutant testes examined, the number of *M34a*-positive cells increased dramatically, both at and displaced away from the apex (Fig. 2h). *Drosophila* germ cells contain characteristic spectrin-rich structures: the ball-shaped spectroosome in stem cells and gonialblasts and the branched fusome in spermatogonial and spermatocyte cysts¹¹. In wild-type

testes, cells with spectroosomes appear in a rosette one or two cells deep around the hub (Fig. 2i). *Egfr*^{ts} mutants had many more cells with ball-shaped spectroosomes, both near the tip (Fig. 2j) and displaced down the testis (Fig. 2k). Dumbbell-shaped spectroosomes were also common, as in recently divided stem cells or cystoblasts in the female germline¹². In wild-type testes, germ cells undergo mitotic division only at the apical tip. Stem cells and gonialblasts divide as single cells while interconnected spermatogonia divide in synchrony⁷ (Fig. 2l). *Egfr*^{ts} mutants had single cells in metaphase, consistent with stem cell or gonialblast identity, throughout the testes (Fig. 2m, arrows).

The earliest germ cells accumulating in *Egfr*^{ts} mutant testes resembled stem cells based on expression of the stem cell marker *escargot* (*esg*)¹³. *esg* function is required for maintenance of male germline stem cells (G. Hime and M.F., manuscript in preparation). In wild-type testes, *esg* messenger RNA was detected in somatic hub cells and the adjacent rosette of germline stem cells, but not in gonialblasts or spermatogonia (Fig. 3a). In all *Egfr*^{ts} mutant testes examined (*n* = 36), *esg* mRNA appeared in many individual, paired or small clusters of cells scattered throughout the testes (Fig. 3b–d). In 29 testes, *esg*-positive germ cells also appeared in an enlarged rosette at the tip (Fig. 3b). In seven testes, the hub was displaced from the tip (Fig. 3d) or not detected. The number of *esg*-positive germ cells per testis ranged from twenty to hundreds, proportional (~10%) to the total number of accumulated early germ cells. The increased number of cells with stem cell division behaviour, spectroosomes and marker expression suggested that without *Egfr*⁺ function stem cells may divide symmetrically, producing two daughters that retain stem cell identity. Stem cell characteristics were even maintained in cells displaced away from the hub, outside the normal stem cell niche.

Wild-type *Egfr* function also appeared to restrict mitotic amplification divisions and allow spermatogonia to differentiate (Fig. 1, arrow III). Cytoplasmic *bag-of-marbles* protein (Bam-C)^{14,15} is normally expressed in two-cell cysts to early sixteen-cell cysts (Fig. 4a). In *Egfr*^{ts} mutants, the number of germ cells lacking detectable Bam-C at the tip was increased (Fig. 4b), consistent with expansion of the stem cell and gonialblast populations. In

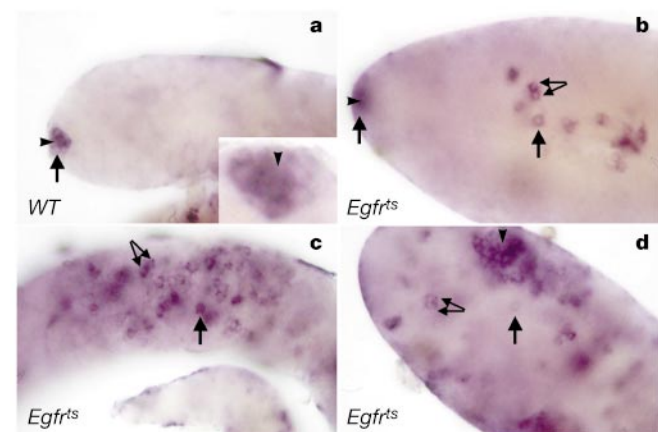


Figure 3 *Egfr*⁺ restricts germline stem cell self-renewal. Germline stem cells (arrows) and somatic hub (arrowheads) stained for *escargot* mRNA. **a**, Control. Hub and surrounding rosette of ~9 germline stem cells (inset, higher magnification of testis tip). **b–d**, *Egfr*^{ts}. Many isolated (arrows), pairs or small clusters (forked arrows) of *esg*-positive cells displaced from apical tip. **b**, Hub at tip. **c**, Distal testis region with many *esg*-positive cells. **d**, Mislocalized hub.

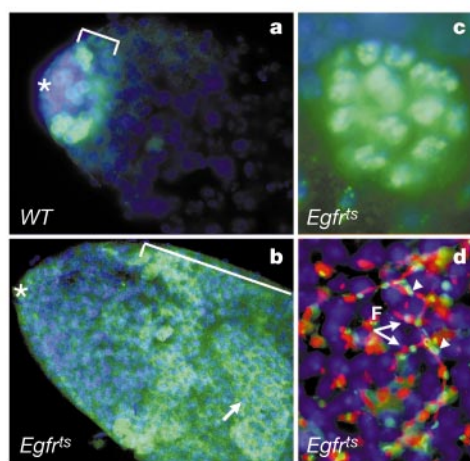


Figure 4 *Egfr*⁺ restricts spermatogonial amplification divisions. **a, b**, Anti-Bam-C (green). DAPI (blue). In wild-type testis (**a**), Bam-C is detected in spermatogonia (bracket) but not stem cells or gonialblasts (asterisk to bracket). In *Egfr*^{ts} testes (**b**) there are expanded Bam-C-negative (asterisk to bracket) and Bam-C-positive populations (bracket). Arrow, cyst of Bam-C-positive cells with more than sixteen spermatogonia. **c**, *Egfr*^{ts} testis stained with anti-phosphorylated-Histone-3 (green). Sixteen spermatogonia undergoing extra round of synchronous mitosis. **d**, *Egfr*^{ts} testis stained for anti-Spectrin (red), anti-phosphotyrosine (green) and DAPI (blue). Large cluster of cells resembling spermatogonia, with branched fusome (arrows) and multiple ring canals (arrowheads).

addition, *Egfr^{ts}* testes had large clusters of Bam-C-positive cells with far more than 16 cells per cyst (Fig. 4b). Consistent with failure of spermatogonia to cease mitotic division, clusters of 16 or more cells in synchronous mitosis (Fig. 4c) and groups of interconnected cells with fusome running through more than 15 ring canals (Fig. 4d) were common. Thus, *Egfr⁺* function is required to restrict self-renewal or proliferation and allow at least two critical differentiation choices during spermatogenesis: the stem cell to gonialblast switch, and the spermatogonia to spermatocyte transition. The large Bam-C-positive cysts may originate from cells that made the stem cell to gonialblast transition and initiated spermatogonial divisions before depletion of *Egfr^{ts}* function.

Egfr function is not required in the germline to restrict germline stem cell self-renewal or allow differentiation of gonialblasts, spermatogonia or spermatocytes. Germline clones homozygous for an *Egfr* null or *Egfr^{ts}* allele produced cysts with 16 differentiating spermatocytes (Fig. 5a). None of the homozygous *Egfr* mutant germline clones ($n = 86$) showed overproliferation of early germ cells. Somatic cyst cell clones homozygous mutant for *Egfr* were not recovered, indicating that loss of *Egfr* function in individual cyst progenitors may cause cell lethality or block differentiation.

Normal differentiation of *Egfr* mutant germline clones indicated *Egfr⁺* acts in somatic cells to restrict germline stem cell self-renewal and allow differentiation. Parallel analysis of the *Egfr* downstream

effector D-Raf¹⁶ identified cyst cells as the critical somatic cells involved¹⁷. Consistent with activation of *Egfr* signalling in cyst cells, wild-type cyst cells associated with early germline had high levels of activated MAP-kinase¹⁸ (Fig. 5b). In addition, expression of *vein-lacZ* was detected in both early and late cyst cells (Fig. 5c). Expression of *vein*, an *Egfr* ligand, is a downstream target of *Egfr* activation in several tissues^{19–21}. *Egfr* may be activated in cyst cells in response to a germline-derived ligand: an enhancer trap in the *Egfr* ligand *spitz* was expressed in germ cells (Fig. 5d). Expression of *spitz* in male germ cells may activate *Egfr* on somatic cyst cells, much as expression of an oocyte-derived ligand activates *Egfr* in overlying somatic follicle cells during oogenesis, inducing *vein* expression and eventual patterning of the egg¹⁹.

As *Egfr⁺* appears to act in somatic cells to allow germline differentiation, we assayed effects on hub and cyst cells in *Egfr^{ts}* mutants. The hub was present in most *Egfr^{ts}* testes (Figs 2d, 3b, 5f). Staining for a nuclear marker²² normally in paired cyst cells surrounding 16 spermatogonia also revealed pairs in the mutant. However, in distal regions of *Egfr^{ts}* testes, the paired cyst cells were often associated with more than 16 germ cells (Fig. 5f, inset). Early cyst cells often formed improper associations with early germ cells in *Egfr^{ts}* testes. In general, accumulation of early germ cells was not associated with a corresponding increase in the total number of cyst progenitor or early cyst cells. In addition, often more than two cyst cells surrounded a mass of early germ cells (observed in 17/50 *Egfr^{ts}* testes) (compare Fig. 5e and f). In one case, we observed an isolated cyst of germ cells with over 15 cyst cell nuclei. Cyst cells surrounding a germline mass might continue to divide rather than differentiate. Alternatively, multiple cyst cells may be recruited during cyst formation. In either case, disruption of the normal association between cyst cells and germ cells could allow proliferation of early germ cells at the expense of differentiation.

Signals from somatic support cells are suggested to specify germline stem cell self-renewal and prevent differentiation in *Caenorhabditis elegans*²³, *Drosophila* oogenesis^{6,14,24,25}, and mammalian spermatogenesis²⁶. Our results suggest an additional role: extrinsic signals from somatic support cells restrict germline stem cell self-renewal and ensure that one stem cell daughter adopts the gonialblast fate and initiates differentiation (Fig. 1, arrow I). We propose that *Egfr⁺* activation in cyst cells sends a signal that prevents self-renewal of stem cell identity by the germ cell they enclose. Without *Egfr⁺* function, the cyst cells do not send the proposed signal and both daughter cells can retain stem cell identity, increasing stem cell number. Thus, support cells act as guardians, reining in the capacity for expansion of the stem cell population through symmetric division and ensuring the correct balance between self-renewal and differentiation.

Mammalian haematopoietic and spermatogonial stem cell populations markedly expand after transplantation^{3,5}. In the *Drosophila* female germline, stem cells can divide symmetrically to replace lost neighbouring stem cells⁶. We propose that, as in the male germline, guardian mechanisms may normally control capacity for extensive stem cell self-renewal in other stem cell systems. □

Methods

Egfr^{ts} mutants

Egfr^{tsla}/Egfr^{CO} trans-heterozygotes^{10,27} were raised at 18 °C, shifted as adults to 29 °C, and aged for one week. Controls were heterozygous siblings.

Cytology

Whole testes stained²⁸ with 1:10 anti-FasciclinIII (Goodman), 1:10,000 anti-Vasa (Lehmann), 1:1,000 C17.9C6 anti-Notch (Artavanis-Tsakonas), 1:200 anti-alpha-Spectrin (Branton), 1:100 anti-phosphorylated-Histone-H3 or anti-phosphotyrosine (Upstate Biotechnology), 1:1,000 anti-Bam-C (McKearin), 1:1,000 anti-Eyes-Absent (Bonini), 1:200 FITC- or TRITC-conjugated anti-IgG (Jackson Immuno Research), and 4,6-diamidino-2-phenylindole (DAPI, Molecular Probes). β -Galactosidase activity from *P{PZ}spi 01068* (Shilo), *P{PZ}vn 10567* (Simcox), *M34a*, *600* and *I72* (ref. 9) detected with X-Gal. For activated-MAPK, testes dissected in 10 mM Tris-Cl pH 6.8, 180 mM KCl,

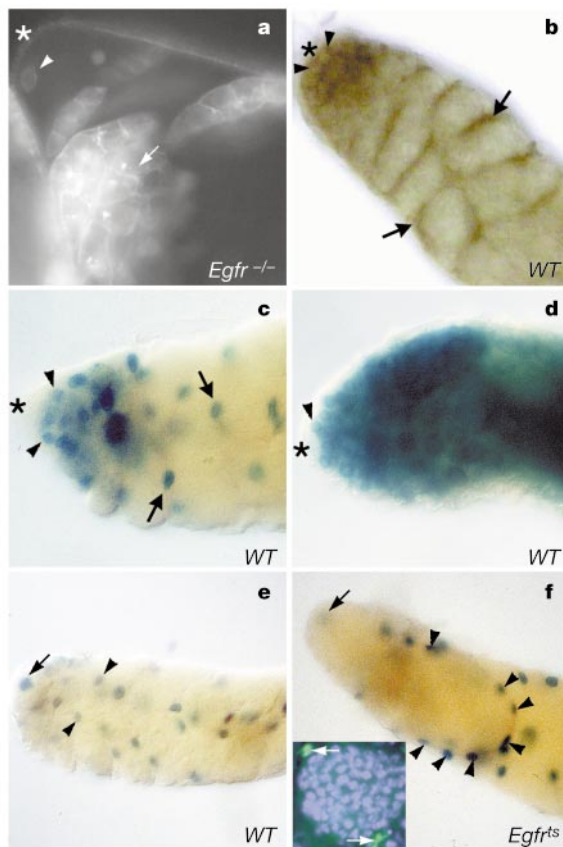


Figure 5 *Egfr⁺* activity in somatic cells regulates germline stem cell differentiation. Asterisk, testes apex. **a**, *Egfr^{ts}* homozygous germline clone showing stem cell (arrowhead), spermatogonia (short arrow) and 16 spermatocytes (arrow). **b**, Wild type. Active MAP-kinase in cyst cells neighbouring germline stem cells (arrowheads) and spermatocytes (arrows). **c**, *vein* enhancer trap in cyst cells. **d**, *spitz* enhancer trap in germline. **e**, **f**, *600* enhancer trap in hub (arrow) and cyst cells (arrowheads) in wild-type (**e**) and *Egfr^{ts}* (**f**) testes. Multiple cyst cells are associated with many germ cells. Inset, *Egfr^{ts}*. More than sixteen germ cells (DAPI, blue) with two Eya-positive cyst cell nuclei (green, arrows).

50 mM NaF, 1 mM NaVO₄, 10 mM β -glycerophosphate were immuno-stained using 1:250 anti-dp-ERK (Sigma M8159), 1:2,000 biotinylated-anti-mouse (Vector), and 1:1,000 extravidin-HRP (Sigma). *In situ* hybridizations used digoxigenin-UTP riboprobes from *esg* complementary DNA (Hayashi).

Mosaic analyses

Marked germline clones homozygous for *Egfr*^{CO} (shown) or *Egfr*^{ts} induced in adults using MARCM with 37 °C heat-shock for one hour²⁹. Full genotype: *UAS-mCD8-GFP;hs-FLP/Y;FRT^{UAS},top^{ts}/FRT^{UAS},tubP-GAL80;tubP-GAL4/+*. Flies aged at 25 °C for 6 days scored for germline phenotypes and GFP expression by phase-contrast and fluorescence microscopy. 50 *Egfr*^{CO} testes (*n* = 182) and 36 *Egfr*^{ts} testes (*n* = 93) had GFP-germline cells: none showed germline overproliferation. *Egfr* mutant GFP-cyst progenitor clones were not detected (*n* = 275), compared with 35 in controls (*n* = 60).

Received 4 July; accepted 4 August 2000.

1. Loeffler, M. & Potten, C. S. In *Stem Cells* (ed. Potten, C. S.) 1–27 (Academic, London, 1997).
2. Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
3. Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E. & Weissman, I. L. Identification of a lineage of multipotent hematopoietic progenitors. *Development* **124**, 1929–1939 (1997).
4. Green, H. Cultured cells for the treatment of disease. *Sci. Am.* **265**, 96–102 (1991).
5. Parreira, G. G. *et al.* Development of germ cell transplants in mice. *Biol. Reprod.* **59**, 1360–1370 (1998).
6. Xie, T. & Spradling, A. C. *decapentaplegic* is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251–260 (1998).
7. Hardy, R., Tokuyasu, T., Lindsley, D. & Garavito, M. The germinal proliferation center in the testis of *Drosophila melanogaster*. *J. Ultrastruct. Res.* **69**, 180–190 (1979).
8. Lindsley, D. & Tokuyasu, K. T. In *Genetics and Biology of Drosophila* (eds Ashburner, M. & Wright, T. R.) 225–294 (Academic, New York, 1980).
9. Gönczy, P. Towards a molecular genetic analysis of spermatogenesis in *Drosophila*. Thesis, Rockefeller Univ., New York (1995).
10. Kumar, J. P. *et al.* Dissecting the roles of the *Drosophila* EGF receptor in eye development and MAP kinase activation. *Development* **125**, 3875–3885 (1998).
11. Lin, H., Yue, L. & Spradling, A. S. The *Drosophila* fusome, a germline specific organelle, contains membrane skeletal proteins and functions in cyst formation. *Development* **120**, 947–956 (1994).
12. de Cuevas, M. & Spradling, A. C. Morphogenesis of the *Drosophila* fusome and its implications for oocyte specification. *Development* **125**, 2781–2789 (1998).
13. Whiteley, M., Noguchi, P. D., Sensabaugh, S. M., Odenwald, W. F. & Kassisi, J. A. The *Drosophila* gene *escargot* encodes a zinc finger motif found in snail-related genes. *Mech. Dev.* **36**, 117–127 (1992).
14. McKearin, D. & Ohlstein, B. A role for the *Drosophila* Bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* **121**, 2937–2947 (1995).
15. Gönczy, P., Matunis, E. & DiNardo, S. *bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development* **124**, 4361–4371 (1997).
16. Brand, A. H. & Perrimon, N. Raf acts downstream of the EGF receptor to determine dorsoventral polarity during *Drosophila* oogenesis. *Genes Dev.* **8**, 629–639 (1994).
17. Tran, J., Brenner, T. & DiNardo, S. Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis. *Nature* **407**, 754–757 (2000).
18. Gabay, L., Seger, R. & Shilo, B. Z. MAP kinase in *in situ* activation atlas during *Drosophila* embryogenesis. *Development* **124**, 3535–3541 (1997).
19. Wasserman, J. D. & Freeman, M. An autoregulatory cascade of EGF receptor signaling patterns the *Drosophila* egg. *Cell* **95**, 355–364 (1998).
20. Wessells, R. J., Grumblin, G., Donaldson, T., Wang, S. H. & Simcox, A. Tissue-specific regulation of vein/EGF receptor signaling in *Drosophila*. *Dev. Biol.* **216**, 243–259 (1999).
21. Golembo, M., Yarnitzky, T., Volk, T. & Shilo, B. Z. Vein expression is induced by the EGF receptor pathway to provide a positive feedback loop in patterning the *Drosophila* embryonic ventral ectoderm. *Genes Dev.* **13**, 158–162 (1999).
22. Bonini, N. M., Leiserson, W. M. & Benzer, S. The eyes absent gene: genetic control of cell survival and differentiation in the developing *Drosophila* eye. *Cell* **72**, 379–395 (1993).
23. Kimble, J. & White, J. G. On the control of germ cell development in *Caenorhabditis elegans*. *Dev. Biol.* **81**, 208–219 (1981).
24. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
25. King, J. F. & Lin, H. Somatic signaling mediated by *fs(1)Yb* is essential for germline stem cell maintenance during *Drosophila* oogenesis. *Development* **126**, 1833–1844 (1999).
26. Meng, X. *et al.* Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489–1493 (2000).
27. Clifford, R. J. & Schubach, T. Coordinately and differentially mutable activities of *torpedo*, the *Drosophila melanogaster* homolog of the vertebrate EGF receptor gene. *Genetics* **123**, 771–787 (1989).
28. Hime, G. R., Brill, J. T. & Fuller, M. T. Assembly of ring canals in the male germ line from structural components of the contractile ring. *J. Cell Sci.* **109**, 2779–2788 (1996).
29. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).

Acknowledgements

We thank J. Kumar and K. Moses for the *Egfr*^{ts} allele; S. DiNardo, D. Traver, S. Kim and Fuller lab members for communication of unpublished data and helpful comments; and the Howard Hughes Medical Institute (A.A.K.) and the National Institutes of Health (M.T.F.) for financial support.

Correspondence and requests for materials should be addressed to M.F. (e-mail: fuller@cmgm.stanford.edu).

Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis

John Tran[†], Tamara J. Brenner[†] & Stephen DiNardo

Department of Cell & Developmental Biology, University of Pennsylvania Medical Center, 1215 BRB II/III, 421 Curie Blvd, Philadelphia, Pennsylvania 19104-6058, USA

[†] These authors contributed equally to this work.

Stem cells divide both to produce new stem cells and to generate daughter cells that can differentiate¹. The underlying mechanisms are not well understood, but conceptually are of two kinds². Intrinsic mechanisms may control the unequal partitioning of determinants leading to asymmetric cell divisions that yield one stem cell and one differentiated daughter cell. Alternatively, extrinsic mechanisms, involving stromal cell signals, could cause daughter cells that remain in their proper niche to stay stem cells, whereas daughter cells that leave this niche differentiate. Here we use *Drosophila* spermatogenesis as a model stem cell system³ to show that there are excess stem cells and gonialblasts in testes that are deficient for Raf activity. In addition, the germline stem cell population remains active for a longer fraction of lifespan than in wild type. Finally, *raf* is required in somatic cells that surround germ cells. We conclude that a cell-extrinsic mechanism regulates germline stem cell behaviour.

The testis proliferation centre in *Drosophila melanogaster* includes the germline and somatic stem cells that maintain spermatogenesis (Fig. 1a)^{3,4}. As a germline stem cell divides, one daughter becomes a gonialblast, while the other remains a stem cell. To

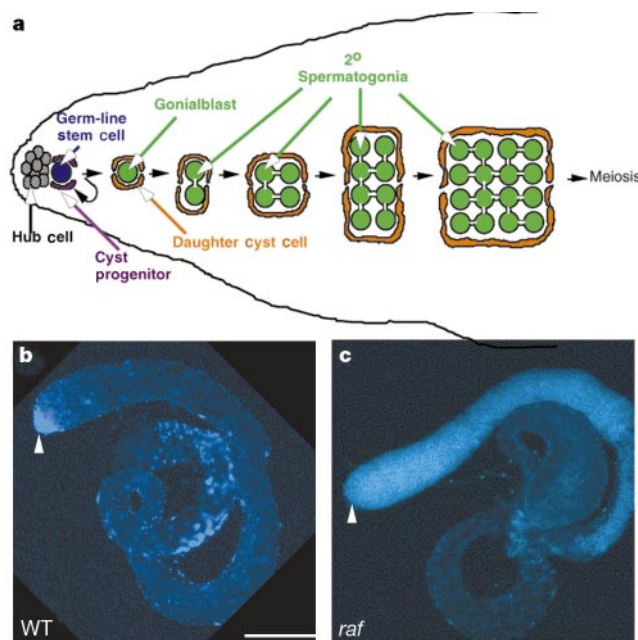


Figure 1 *raf* function restricts early germ cell number. **a**, Testis apex, where 5–9 germline stem cells (one shown for clarity) and approximately twice as many somatic cyst progenitor cells are anchored around somatic hub cells. The testis proliferation centre, comprising hub, germline and somatic stem cells, gonialblasts and 2° spermatogonia (which undergo incomplete cytokinesis; white bars), is restricted to the tip as shown by bright DNA stain (**b**, arrowhead). Spermatocyte and postmeiotic spermatids, which fluoresce more weakly owing to chromatin reorganization, fill the remainder of the testis. **c**, *raf*-deficient testis on day 15. Scale bar, 100 μ m.

amplify the germline population, each gonialblast executes four divisions as 2° spermatogonia, which exit the mitotic cycle and enter a meiotic and differentiation programme as a clone of 16 spermatocytes. The gonialblast and its progeny are encysted by somatic cells derived from cyst progenitor stem cells³. We previously characterized mutants affecting 2° spermatogonia^{5,6}, and have begun to test signal transduction pathways for possible involvement at earlier decision points in this germline stem cell lineage. Null mutations in *raf*, encoding a serine/threonine kinase involved in receptor tyrosine kinase pathways^{7,8}, are lethal in larvae, but such flies carrying a heat shock (HS)-*Raf* transgene can survive to fertile adults by using daily heatshocks⁹. By withdrawing heat shock when adults eclose, animals become progressively *raf* deficient as the protein decays. Testes from such hypomorphic *raf*-deficient males 5 days after eclosion were indistinguishable from controls. By day 7, however, 43 of 44 *raf* mutant testes showed great proliferation centre expansion. The increase in cell number is due to excess, early stage germ cells, and continues such that on day 15, *raf*-deficient testes are filled with

these cells at the expense of post-mitotic cells (compare Fig. 1b and c).

To identify the earliest defect in germ-cell progression, we examined the fusome, a membrane and cytoskeletal organelle specific to the germline that is spheroid throughout stem cells and gonialblasts, but branches extensively throughout interconnected 2° spermatogonia (Fig. 2a)^{10–12}. In *raf*-deficient testes, fusome structure is not normal. Unbranched fusomes are found in many cells (Fig. 2b), even those located some distance from the hub, where, in wild-type testes, only branched fusomes interconnecting 2° spermatogonia are found. Although branched fusomes do appear in *raf*-deficient testes, unlike wild-type testes, these fusomes usually appear only after an intervening region containing many excess germ cells that have only spheroid fusomes. These excess germ cells do not result from an increased frequency of germline stem cell divisions, as the M-phase index for the tier of cells adjacent to the hub in *raf*-deficient testes was almost identical to that in the wild type (about 3 M phases in 14 testes on day 13).

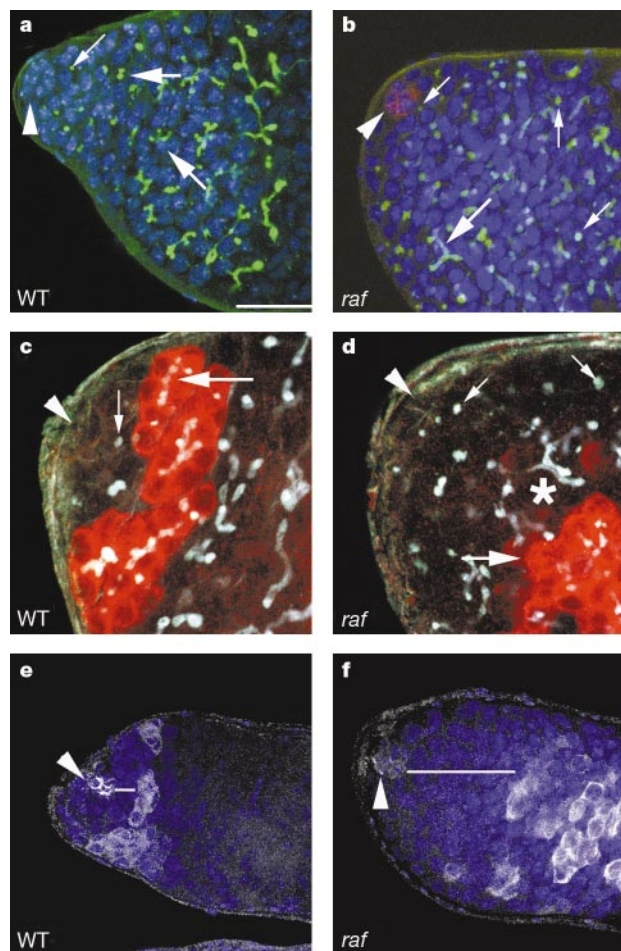


Figure 2 Early stage germ cells accumulate in *raf*-deficient testes. Confocal projection of half testis depth; DNA, purple (**a, b, e, f**). **a**, Wild-type (WT) testis fusome (mAb1b1) is spherical in stem cells (small arrow) and gonialblasts, which are located only near the hub (arrowhead). Fusome branches in 2° spermatogonia (larger arrows). **b**, *raf*-deficient testis, day 8. Spherical fusomes (small arrows) are found both next to and far away from the hub (arrowhead; anti-Fasciclin (Fasc) III). Fusome branching is detected (larger arrow), more so farther from the hub (not shown). **c**, WT testis, day 6; Bam-C (red) is expressed only in 2° spermatogonial cells, four groups of which are shown, all having branched fusomes (large arrow; white). Cells to the right of this are maturing spermatocytes, in which Bam-C protein has decayed, and all of which possess fusomes branching into and out of the focal plane. The hub is just below the focal plane (arrowhead). **d**, *raf*-deficient testis, day 6; two of nine cells with unbranched fusomes are indicated (small arrows), none of which

accumulates Bam-C. The fusome is branched in the spermatogonial cluster expressing Bam-C (large arrow). The asterisk shows one of two 2° spermatogonial cysts with interconnected fusomes that either did not accumulate Bam-C, or lost it precociously. The hub is just below focal plane (arrowhead). **e**, WT testis, day 8; only a thin band of cells (line) separate Bam-C-expressing 2° spermatogonial cells (white) from the hub (arrowhead; white). **f**, *raf*-deficient testis, day 8; many (line) germline stem cells and gonialblasts separate Bam-C-expressing 2° spermatogonial cells (white) from the hub (arrowhead; white). Further along the testis, Bam-C-expressing 2° spermatogonia accumulate in cysts of up to 35 cells, as shown by S- or M-phase synchrony (not shown). Interspersed with these clusters are individual cells that cycle in isolation and do not express Bam-C. Scale bar, 20 μ m in **a, b, e, f**; 10 μ m in **c, d**.

We next examined cytoplasmic Bag-of-marbles protein (Bam-C), which first accumulates in 2° spermatogonial cells (Fig. 2c) and is required for their progression into spermatocytes^{6,13}. In the wild type, the total population of germline stem cells and gonialblasts comprises the few non-staining cells between the hub and the first rows of 2° spermatogonia (Fig. 2e). In *raf*-deficient testes, the first Bam-C-expressing germ cells are located much further from the hub (Fig. 2f). Also, most of the intervening non-staining cells contain unbranched fusomes (Fig. 2d), consistent with these cells representing excess gonialblasts and/or germline stem cells.

To verify this, we examined M5-4 and S1-33, markers expressed in hub cells, germline stem cells and gonialblasts, but not in 2° spermatogonia (Fig. 3a)⁶. In *raf*-deficient testes, while hub cells appeared normal (Fig. 2d, 3b), the number of germ cells expressing these markers was greatly increased (Fig. 3b). Furthermore, marker expression persisted for a significantly longer fraction of adult lifespan. For instance, by day 19, all control testes ($n = 14$) had M5-4 marker expression in fewer than three germ cells (Fig. 3c), rather than the average 5 to 9 germline stem cells plus associated gonialblasts contained in testes from young adults¹⁴. In contrast, only 17% of *raf*-deficient testes ($n = 23$) on day 19 showed loss of germ cell expression, whereas 48% maintained expression at least equivalent to that of day 1, and a further 35% still showed increased expression (Fig. 3d). These data suggest that germline stem cells and gonialblasts remain active for a significantly longer period than in the wild type. We tested this further by directly counting the number of cycling germline stem cells, which we judged to be cells that were located adjacent to the hub and contained a spherical fusome and

nuclear anillin, a late interphase marker. Whereas wild-type testes averaged 3.1 late interphase germline stem cells per testis ($n = 10$ testes; standard deviation, s.d. = 1.0) on day 1, on day 19 aged cohorts had only 0.8 ($n = 10$; s.d. = 1.0; $P < 10^{-4}$; Fig. 3e). This suggests an age-induced quiescence of the germline stem cell population. In contrast to this, in *raf*-deficient testes late interphase germline stem cell numbers were maintained over 19 days (2.5 per testis on day 1, $n = 10$, s.d. 1.0; 2.8 per testis on day 19, $n = 10$, s.d. = 1.7; Fig. 3f) suggesting that germline stem cells as a population remain active for a longer period than in wild-type.

To determine whether *raf* function is required in the germ line, which exhibits the phenotypes described above, or in the surrounding somatic cells, we generated *raf* null mutant clones. Persistent germline clones indicate the existence of a *raf* null germline stem cell^{3,6}. In all cases (34% of testes; $n = 50$), progression through spermatogenesis is normal as judged by groups of 16 mutant germ cells that are morphologically indistinguishable from surrounding wild-type spermatocytes (Fig. 4a). Thus, *raf* function is dispensable in germ cells. In contrast, we recovered no cyst cell clones, suggesting that *raf* is necessary for viability or proliferation of cyst progenitor cells. As our previous analysis was under hypomorphic conditions for *raf*, we repeated the mosaic analysis, introducing a HS-Raf transgene to provide a basal level of *raf* function so that *raf*-deficient cyst cells might survive. Persistent *raf*-deficient cyst cell clones (5% of testes; $n = 204$), indicate the existence of a *raf*-deficient cyst progenitor cell. Such testes had excess early stage, *raf*+ germ cells (Fig. 4b, and inset). Thus, *raf* is required in the cyst cell lineage. In addition, in *raf*-deficient testes two cyst cell markers normally expressed in later-stage cyst cells, LacZ600 (Fig. 4c)¹⁵ and Eyes absent (not shown), were now expressed prematurely in somatic cells adjacent to the hub, probably the cyst progenitor cells (Fig. 4d). This shows that *raf* function is required in somatic cells surrounding germline stem cells.

Germline stem cell divisions lead to a distinction between a self-

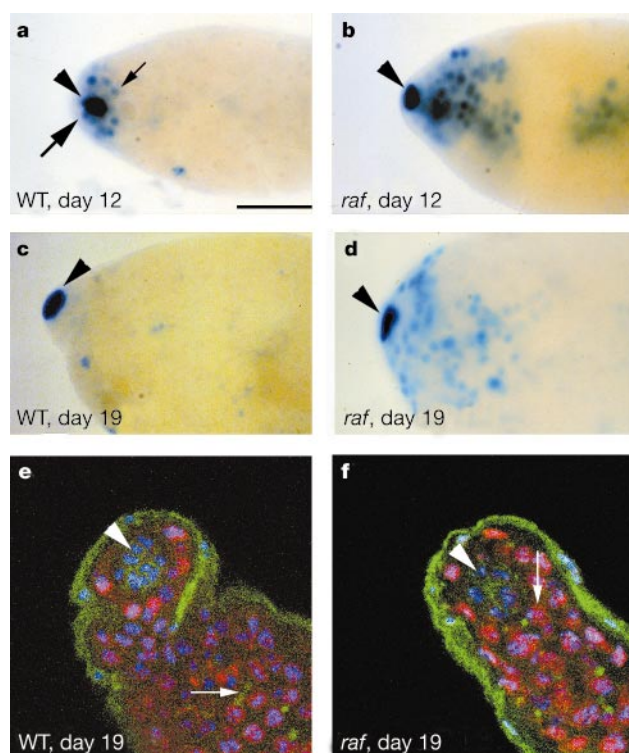


Figure 3 Prolonged activity for the stem cell population in *raf*-deficient testes. **a**, WT testis, X-gal activity; M5-4 expressed in the hub (arrowhead), germline stem cells (large arrow), and gonialblasts (smaller arrow). **b**, *raf*-deficient testis; M5-4 is expressed in many cells near hub (arrowhead). **c**, WT testis, day 19; expression solely in hub. **d**, *raf*-deficient testis, day 19. **e**, WT testis, day 19; branched fusomes (green) are visible further along the testis (arrow), but there are no cells with spherical fusomes adjacent to hub (arrowhead). **f**, *raf*-deficient testis, day 19; an example of late-interphase (nuclear Anillin, red) germline stem cell (arrow), as indicated by the presence of a spherical fusome (green) in the cell adjacent to the hub (arrowhead). Scale bar, 40 μ m.

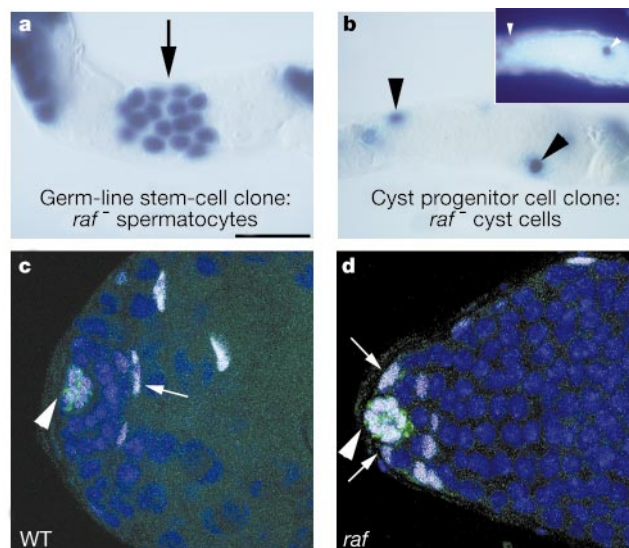


Figure 4 *raf* is required in cyst progenitor cells. **a**, Normal spermatocytes (β -galactosidase-positive, arrow) produced 10 days after induction of a *raf* germline stem cell clone. Similar data were obtained for null mutations in *egfr* and *ras1*. **b**, Deficiency of *raf* in somatic cyst cells (arrowheads) leads to excess early stage germ cells (inset; DNA stain). Glutaraldehyde fixation for LacZ activity meant that we could not assay germ cells for stem or gonial cell markers. **c**, WT testis, day 18. LacZ600 (white; anti- β -galactosidase) is expressed in the hub (arrowhead; green) and in later cyst cells (arrow). **d**, *raf*-deficient testis, day 18. LacZ600, expressed in cyst progenitor cells (arrows) adjacent to the hub (arrowhead); this is visible before day 9. Scale bar, 100 μ m in **a**, **b** (125 μ m in inset); 20 μ m in **c**, **d**.

renewing daughter cell and a sister cell committed to differentiation as a gonialblast. Our data suggest that a somatic signal influences this decision by limiting the self-renewing potential. We cannot yet say how this potential is encoded. Our hypothesis is that in *raf*-deficient testes, where cyst progenitor cell identity is disturbed, the signal is lost, and excess stem cell potential is produced. Upon division, both daughters of the germline stem cell inherit some stem cell character. At steady state, this increases the number of cells that become gonialblasts, and somehow prolongs the active state of the stem cell population. Thus, a somatic cell defect leads to a tumour in the germline stem cell lineage, which suggests that some tumours of progenitor cell populations could be initiated by genetic lesions in support cells, rather than in the tumorous cells themselves. In the testis, the Raf-dependent signal may be delivered by cyst progenitor cells, or their cyst-cell daughters. Our mosaic analysis shows that depletion of Raf from just one of the two cyst progenitor cells surrounding a germline stem cell causes a defect. This may indicate a dose effect, where one heterozygous somatic cell is not sufficient to allow normal signalling. It is likely that the signal transducer Raf is engaged, owing to activation of the Epidermal growth factor receptor pathway in somatic cells¹⁶. We note that in *raf*-deficient testes, differentiation of 2° spermatogonia is blocked as they do not transit to the spermatocyte stage. We believe this is a secondary effect owing to the defective cyst lineage, as we previously showed that a cyst cell signal governs this later transition from 2° spermatogonia to spermatocytes⁵.

Somatic signals have been postulated to affect germline stem cell behaviour in the *Drosophila* ovary^{17,18}. However, the characterized signals are necessary to maintain germline stem cells, rather than restrict their self-renewing potential, as we find here. Additionally, *raf*-deficient ovaries exhibit no increases in early stage germ cells¹⁹. Thus, despite the superficial similarities of early germ cell development in ovary and testis, oogenesis and spermatogenesis are emerging as complementary systems from which different principles of stem cell regulation will emerge. □

Methods

The *raf*¹¹⁻²⁹ or *raf*^{400B6}, HS-Raf/+ larvae were heat shocked for 1 hour at 38 °C, daily. At eclosion (day 1), heat shocking was halted, and flies were aged with wild-type FM6/Y siblings. Stains were carried out as in refs 3,11 and 15. Figure 3e, f shows wild-type and mutant cohorts that were aged and processed using Anti-Anillin for cell-cycle stage²⁰, Anti-Fasc III and Anti-1B1. Cells adjacent to the hub containing dot fusomes and exhibiting nuclear anillin accumulation were scored as late interphase germline stem cells. Wild-type and *raf*-deficient testes differed significantly at day 19 ($P < 0.003$). Figure 4a, b shows clones induced by a 1-hour heat shock at 38 °C in day 1 *raf*^{400B6} Act> DRaf>nuc-lacZ / Y; HS-Flp/+ or *raf*^{400B6} Act5C>DRaf>nuc-lacZ / Y; HS-Raf /+; HS-Flp/+ males that were fixed and stained on day 10. The DRaf+ flip-out is described in ref. 21. The continued production of persistent marked clones demonstrates the marking of a stem cell⁶.

Received 2 June; accepted 28 July 2000.

- Hall, P. A. & Watt, F. M. Stem cells: the generation and maintenance of cellular diversity. *Development* **106**, 619–633 (1989).
- Horvitz, R. H. & Herskowitz, I. Mechanisms of asymmetric cell division: two B's or not two B's, that is the question. *Cell* **68**, 237–255 (1992).
- Gönczy, P. & DiNardo, S. The germ line regulates somatic cyst cell proliferation and fate during *Drosophila* spermatogenesis. *Development* **122**, 2437–2447 (1996).
- Fuller, M. T. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 71–147 (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1993).
- Matunis, E. *et al.* *punt* and *schurri* regulate a somatically derived signal that restricts proliferation of committed progenitors in the germline. *Development* **124**, 4383–4391 (1997).
- Gönczy, P., Matunis, E. & DiNardo, S. *bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development* **124**, 4361–4371 (1997).
- Ambrosio, L., Mahowald, A. P. & Perrimon, N. Requirement of the *Drosophila raf* homologue for *torso* function. *Nature* **342**, 288–291 (1989).
- Dickson, B. *et al.* Raf functions downstream of Ras1 in the Sevenless signal transduction pathway. *Nature* **360**, 600–603 (1992).
- Lee, T., Feig, L. & Montell, D. J. Two distinct roles for Ras in a developmentally regulated cell migration. *Development* **122**, 409–418 (1996).
- Lin, H., Yue, L. & Spradling, A. C. The *Drosophila* fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. *Development* **120**, 947–956 (1994).
- de Cuevas, M. & Spradling, A. C. Morphogenesis of the *Drosophila* fusome and its implications for oocyte specification. *Development* **125**, 2781–2789 (1998).

- Hime, G. R., Brill, J. A. & Fuller, M. T. Assembly of ring canals in the male germ line from structural components of the contractile ring. *J. Cell Sci.* **109**, 2779–88 (1996).
- McKearin, D. & Ohlstein, B. A role for the *Drosophila* bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* **121**, 2937–2947 (1995).
- Hardy, R. W. *et al.* The germinal proliferation center in the testis of *Drosophila melanogaster*. *J. Ultrastruct. Res.* **69**, 180–190 (1979).
- Gönczy, P., Viswanathan, S. & DiNardo, S. Probing spermatogenesis in *Drosophila* with P-element enhancer detectors. *Development* **114**, 89–98 (1992).
- Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**, 750–754 (2000).
- Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
- Xie, T. & Spradling, A. C. *decapentaplegic* is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251–260 (1998).
- Tran, T. *Genetic Analysis of Germ Line Stem Cell Regulation in Drosophila* 122 (Rockefeller Univ., New York, 1998).
- Field, C. M. & Alberts, B. M. Anillin, a contractile ring protein that cycles from the nucleus to the cell cortex. *J. Cell Biol.* **131**, 165–178 (1995).
- Struhl, G. & Basler, K. Organizing activity of wingless protein in *Drosophila*. *Cell* **72**, 527–540 (1993).

Acknowledgements

We thank our laboratory staff, B. Calvi, A. Kiger, M. Fuller, E. Matunis and S. Wasserman for comments and suggestions. Fly stocks and reagents were contributed by the Bonini, Lipschitz, McKearin, Struhl, Wasserman and Xu labs, as well as the Bloomington Stock center and Iowa Hybridoma bank. The NSF and the NIH supports our work.

Correspondence and requests for materials should be addressed to S.D. (e-mail: sdinardo@mail.med.upenn.edu).

The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*

John I. Glass*†, Elliot J. Lefkowitz*, Jennifer S. Glass*†, Cheryl R. Heiner*‡, Elson Y. Chen*‡ & Gail H. Cassell* †

* Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

† Infectious Diseases Research and Clinical Investigation, Eli Lilly and Company, Indianapolis, Indiana 46285, USA

‡ Advanced Center for Genomic Technology, Perkin-Elmer Corporation, Foster City, California, 94404 USA

The comparison of the genomes of two very closely related human mucosal pathogens, *Mycoplasma genitalium* and *Mycoplasma pneumoniae*, has helped define the essential functions of a self-replicating minimal cell, as well as what constitutes a mycoplasma. Here we report the complete sequence of a more distant phylogenetic relative of those bacteria, *Ureaplasma urealyticum* (*parvum* biovar), which is also a mucosal pathogen of humans. It is the third mycoplasma to be sequenced, and has the smallest sequenced prokaryotic genome except for *M. genitalium*. Although the *U. urealyticum* genome is similar to the two sequenced mycoplasma genomes^{1,2}, features make this organism unique among mycoplasmas and all bacteria. Almost all ATP synthesis is the result of urea hydrolysis, which generates an energy-producing electrochemical gradient. Some highly conserved eubacterial enzymes appear not to be encoded by *U. urealyticum*, including the cell-division protein FtsZ, chaperonins GroES and GroEL, and ribonucleoside-diphosphate reductase. *U. urealyticum* has six closely related iron transporters, which apparently arose through gene duplication, suggesting that it has a kind of respiration system not present in other small genome bacteria. The genome is

§ Present address: Celera Genomics, PE Corporation, Foster City, California 94404, USA

only 25.5% G+C in nucleotide content, and the G+C content of individual genes may predict how essential those genes are to ureaplasma survival.

Ureaplasma urealyticum, a common commensal of the urogenital tract of humans, is gaining recognition as an important opportunistic pathogen during pregnancy. It is the most common organism isolated from the chorioamnion, even in the presence of intact fetal membranes. It is associated with inflammation, premature spontaneous delivery, septicaemia, meningitis and pneumonia in newborn infants³. This bacterium is a member of the class Mollicutes, commonly referred to as mycoplasmas. The Mollicutes are the smallest-known free-living microorganisms. Although they evolved from Gram-positive ancestors, mycoplasmas lack a cell wall, which is their most distinguishing feature. The species that parasitizes humans, *U. urealyticum*, is further subtyped into 14 serovars;

however, recent molecular characterization of the *U. urealyticum* serovars has led to an effort to reclassify them into two species⁴. *U. urealyticum* would include the ten large genome serovars (0.88–1.2 megabase pairs (Mbp)); *U. parvum* would include the small genome serovars 1, 3, 6 and 14 (0.75–0.76 Mbp). We have sequenced the genome of an isolate of *U. urealyticum* serovar 3 (or *U. parvum* in the proposed revised classification), the serovar most commonly isolated from humans⁵.

The genome of *U. urealyticum* serovar 3 is a circular chromosome comprising 751,719 bp. This is smaller than any other sequenced microbial genome except *M. genitalium*¹. The G+C content is 25.5%. The genome contains 613 predicted protein-coding genes and 39 genes that code for RNAs. The genes comprise 93% of the genome. We assigned biological roles to 53% of the protein-coding genes (see Supplementary Information); 19% of the genes are

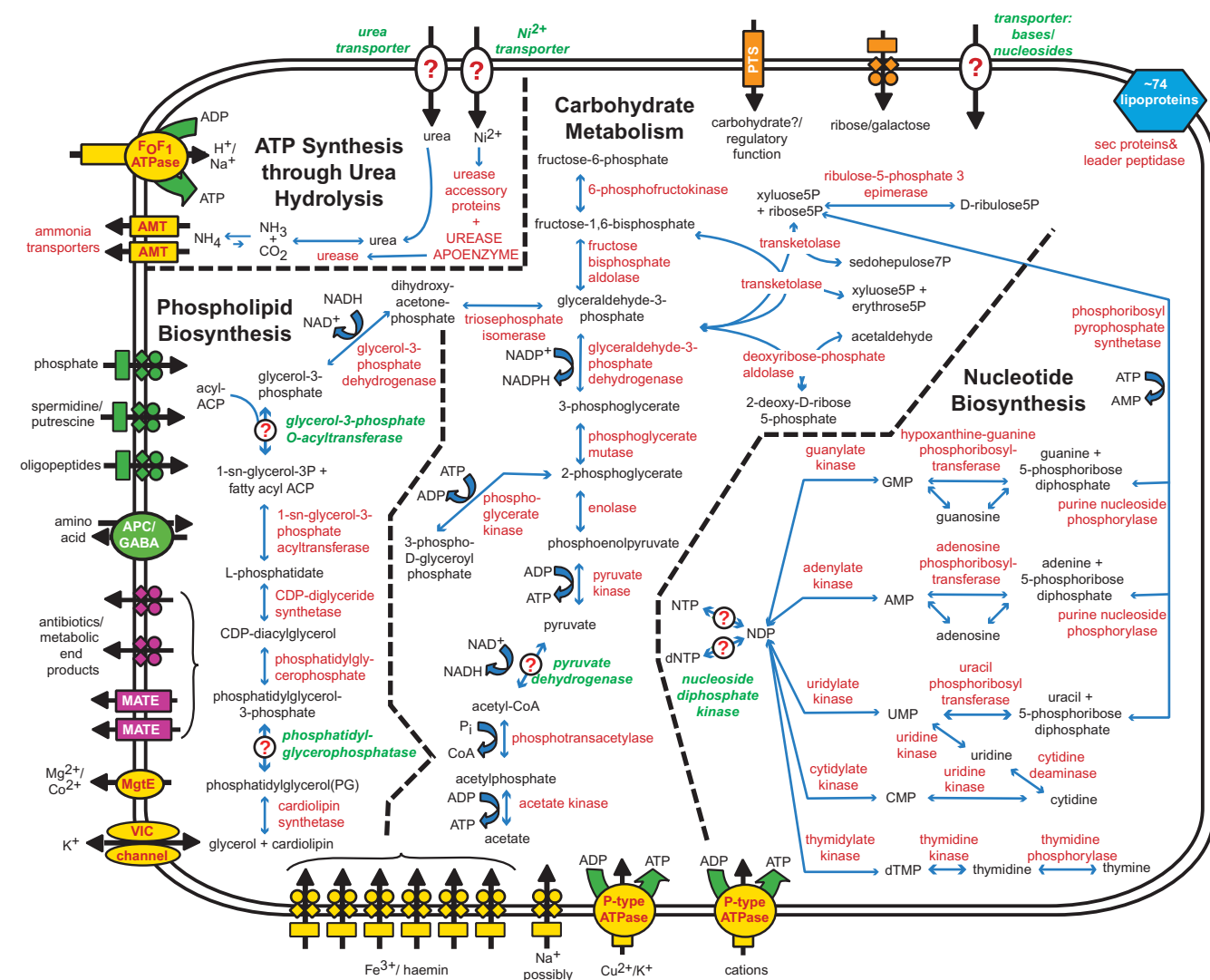


Figure 1 Metabolic pathways and substrate transport in *U. urealyticum*. An integrated view is shown of the transporters and main metabolite elements of an *U. urealyticum* cell, deduced from the set of genes for which we can predict functions. Metabolic products are shown in black and ureaplasma proteins in red. Question marks denote enzymes or transporters not identified that would be necessary to complete pathways, and missing enzyme and transporter names are shown in green italics. Transporters are coloured according to their substrates: yellow, cations; green, anions and amino acids; orange, carbohydrates; purple, multidrug and metabolic end product efflux. Arrows indicate the direction of substrate transport. Transporters are (1) an F-type ATPase; (2) two Amt ammonium transporters; (3) CopA, a copper-importing P-type ATPase, and PaCL, a cation transport P-type ATPase; (4) a K⁺ channel from the voltage-gated ion channel (VIC)

superfamily; (5) a Mg²⁺ MgtE transporter; (6) XasA a glutamate:GABA antiporter from the amino-acid-polyamine-organocation (APC) transporter superfamily; (7) two multidrug antimicrobial extrusion family transporters (MATE); (8) a *ptsH* element of a phosphoenolpyruvate-dependent sugar phosphotransferase transport system (PTS), which probably has a regulatory instead of transport function as the genome lacks the PTS EI component needed for sugar transport; and (9) a broad array of ABC transporters. The ABC-type transporters are shown with a rectangle for the substrate-binding protein, diamonds for the membrane-spanning permeases and circles for the ATP-binding subunits. In some cases we cannot identify all the components of these transporters. ABC efflux transporters have no substrate-binding domains.

similar to hypothetical genes of unknown function, and 28% are unique hypothetical genes with no significant similarities to putative or demonstrated genes in other organisms. On the basis of the gene distribution on the two strands and on transitions in GC skew, we project that the *U. urealyticum* origin of replication is upstream of *dnaA*, which is gene UU001 in our nomenclature.

U. urealyticum generates 95% of its ATP through the hydrolysis of urea by urease⁵. Since the discovery in 1966 that ureaplasmas metabolize urea, the physiology of this energetic process has been studied to gain an understanding of what appears to be a unique system⁶. The hypothesis is that hydrolysis of urea generates an electrochemical gradient through accumulation of intracellular ammonia/ammonium. The gradient fosters a chemiosmotic potential that generates ATP (Fig. 1).

Studies of *U. urealyticum* energy production predict three enzymatic components in this process: urease; an ammonia/ammonium transporter; and an F_0F_1 -ATPase⁵. The role in this process of the highly potent *U. urealyticum* urease, which is 30–180-fold more efficient than that reported for other bacterial ureases⁷, has been shown^{8,9}. Urease is a three-subunit Ni^{2+} -binding enzyme. The other four *ure* genes code for accessory proteins that supply the functional metalcentre to the apoenzyme. Although a nickel transporter could not be identified, candidates for genes encoding the Ni^{2+} transporter or elements of it include the Mg^{2+} transporter *mgtE* (ref. 10) and two hypothetical genes, UU101 and UU555, which had polyhistidine regions characteristic of Ni^{2+} -binding proteins. There were no proteins similar to known urea transporters in either prokaryotes or eukaryotes.

The intracellular concentration of ammonia generated through urea hydrolysis in *U. urealyticum* has been measured at 21 times the extracellular concentration. This excess suggests efflux of ammonium/ammonia through a saturable uniporter rather than by passive diffusion⁵. We identified a pair of genes that code for ammonia transporters of the Amt family. Although the physiology of Amt family transporters is controversial^{11–13}, the probable mechanism in ureaplasma is energy-independent-facilitated diffusion. The electrochemical gradient created by the export of ammonium presumably results in an ATP molecule generating proton influx through the *U. urealyticum* F-type ATPase.

As predicted by the inhibition of its ATP synthesis by *N,N'*-dicyclohexylcarbodiimide, *U. urealyticum* contains an F_0F_1 -ATPase⁵. This family of enzymes comprises two oligomeric complexes: the F_0 transmembrane proton channel, and the F_1 catalytic complex. The ureaplasma δ subunit of the F_1 complex, which in other organisms is the interface between the F_0 and F_1 subunits and causes the permeability of the transmembrane proton channel F_0 , is present as two distinct peptides. UU134 corresponds to the amino-terminal two-thirds of the consensus eubacterial δ , and UU133 corresponds to the normally more conserved carboxyl terminus. Speculatively, these atypical features may be linked to the unique ATP production pathways of ureaplasmas. The sources of the 5% of ureaplasma ATP production not resulting from urea hydrolysis are most probably substrate phosphorylation⁵ (Fig. 1). Enzymatic studies show that *U. urealyticum* has a limited capacity to metabolize carbohydrates.

We did not find genes for the *de novo* biosynthesis of purines or pyrimidines in *U. urealyticum*. Pathways for synthesis of RNA and DNA precursors are largely complete except for several notable missing enzymes (Fig. 1). There is no recognizable ribonucleoside-diphosphate reductase for conversion of ribonucleosides to deoxyribonucleosides, unlike in *M. genitalium* and *M. pneumoniae*. So *U. urealyticum* must import all its deoxyribonucleosides and/or deoxyribonucleoside precursors, or have a mechanism for converting ribonucleosides to deoxyribonucleosides.

As a result of its very limited biosynthetic capacity, ureaplasmas must import more of the nutrients needed for growth than do most other bacteria. We can identify 28 different *U. urealyticum*

transporters, which represent 9 transporter families (Fig. 1). Despite this diversity of transport systems, there are a number of surprising absences, such as the previously mentioned lack of transporters for bases, nucleotides, nickel and urea. Missing transporters may be found among the 33 hypothetical *U. urealyticum* proteins that have 5 or more predicted transmembrane regions.

Unexpected among the 19 ABC transporters are six different Fe^{3+} and/or haemin transporters. Iron is essential for the function of many enzymes; however, the low solubility of Fe^{3+} at physiological pH forces most organisms to develop special mechanisms to acquire necessary amounts. Phylogenetic analysis of the ten membrane-spanning protein subunits of the *U. urealyticum* iron transporters indicated that they all cluster together with respect to other known bacterial proteins and form two related, but distinct paralogous groups within this *U. urealyticum* gene family. Thus, we think that this paralogous family arose from gene duplication events during ureaplasma evolution. In the apparent absence of the complement of transcriptional regulators found in most other bacteria, we speculate that *U. urealyticum* increased its capacity to import iron by increasing the number of iron transporter genes. This indicates that the availability of iron may be a limiting factor of ureaplasma growth; but this is puzzling. Because mycoplasmas have flavin-terminated respiratory chains, they lack the redox enzymes that normally use iron in the form of cofactors. This is especially true for *U. urealyticum* because none of its predicted enzymes requires iron. The most likely explanation is that it uses iron in respiration.

Iron is associated with the membranes of *Mycoplasma capricolum*¹⁴ and *M. pneumoniae*¹⁵. In these mycoplasmas, the iron may act as a reversible electron carrier in concert with NADH oxidase, or as a component of an iron-containing electron storage protein¹⁵. Unlike the other mycoplasmas, *U. urealyticum* lacks both NADH oxidase activity¹⁶ and an NADH oxidase gene. Even if ureaplasmas use iron in their respiration, the iron use is probably different from that in other mycoplasmas.

Five ureaplasma proteins, urease¹⁷, immunoglobulin- α (IgA) protease¹⁸, phospholipases A and C²⁰, and MBA²⁰, along with the ureaplasma enzymatic machinery for generating hydrogen peroxide have been proposed as virulence factors. The pathogenic effect of urease is caused by its generation of ammonia¹⁷. IgA protease may give ureaplasmas the capacity to invade the upper urogenital tract by degrading a principal component of the mucosal immune system. Ureaplasma phospholipases may be responsible for premature labour by altering prostaglandin biosynthesis¹⁹. Neither an IgA protease nor phospholipase A or C can be identified. These *U. urealyticum* enzymes may have diverged so far from orthologues in other bacteria that they are unrecognizable, or they may have convergently evolved an IgA protease and phospholipases with no recognizable sequence similarity to known enzymes. For instance, UU292 is a conserved hypothetical lipoprotein with a GDSL motif characteristic of some lipases²¹, and UU441 is a unique hypothetical gene that contains a motif found in the subtilase family of serine proteases (the ureaplasma IgA protease is a serine protease¹⁸).

Hydrogen peroxide is the haemolysin of *M. pneumoniae* and other mycoplasmas, which led to the hypothesis that mycoplasma virulence is in part the result of reactive oxygen being secreted in close proximity to host cell membranes²². Unlike *M. pneumoniae*, *U. urealyticum* haemolysin activity is not inhibited by catalase²³. Thus, ureaplasma haemolytic activity may be enzymatic. *U. urealyticum* has two haemolysins. The *hlyC* gene has a corresponding orthologue in *M. pneumoniae*; however, because *M. pneumoniae* haemolysis is attributed to H_2O_2 , it is likely that the principal haemolysin in *U. urealyticum* is *hlyA*. Therefore *hlyA* may be a new ureaplasma virulence factor. Orthologues of the haemolysin *hlyA* have both haemolysin and cytotoxic activity in *Serpulina hyodysenteriae* and several mycobacteria. The haemolysis is thought to be due to pore formation. *Serpulina* and mycobacteria species lacking this gene are non-pathogenic²⁴.

Comparisons of two very closely related bacterial genomes, *M. pneumoniae* and *M. genitalium*, have helped to define the essential functions of a self-replicating minimal cell, as well as what constitutes a mycoplasma^{25,26}. Determining what is a minimal cell was aided by the fact that the gene set of *M. genitalium* was essentially a subset of the gene set of *M. pneumoniae*; however, because the evolutionary divergence of those two mycoplasmas is so recent, the comparison has limited resolution. Phylogenetic analysis shows that *U. urealyticum* is a member of the same clade of the class Mollicutes as *M. pneumoniae* and *M. genitalium*; but it is different enough to make the comparison more informative. Of the 613 *U. urealyticum* protein-coding genes, only 324 are homologous to *M. genitalium* genes and *M. pneumoniae* genes (all *M. genitalium* genes have orthologues in *M. pneumoniae*²⁵). No function can be predicted for 77 of the genes shared by the 3 mycoplasmas. Ten of those conserved hypothetical genes are found only in mycoplasmas (see Supplementary Information or <http://genome.microbio.uab.edu/uu>).

edu/uu).

The evolutionary divergence of *U. urealyticum* from the other mycoplasmas is shown through analysis of the relative gene order of orthologous genes among the three species. The relative gene order in *M. genitalium* and *M. pneumoniae* is largely conserved, although additional genes are interspersed in the larger genome of *M. pneumoniae* (Fig. 2). Similar comparisons of the relative genome locations of *M. genitalium* and *U. urealyticum* orthologues, and *M. pneumoniae* and *U. urealyticum* orthologues show that there is very little conservation of gene order across large fractions of the genomes. The regions of synteny between *U. urealyticum* and *M. genitalium* or *M. pneumoniae* consist, in most cases, of genes that are functionally related, such as the ribosomal proteins or the oligopeptide transport genes.

We can predict a function or cellular location for 76 proteins coded for by *U. urealyticum* that are not in the *M. genitalium* and *M. pneumoniae* genomes. Many of those genes are involved in ATP production by urea hydrolysis, which is unique to ureaplasmas, and in iron acquisition. These differences may be why ureaplasmas grow more robustly *in vitro* than other minimal bacteria, and can replicate in both the lower and upper urogenital tracts. The ureaplasma phenotype may largely be the result of its unique set of energy production and respiration genes.

Because the *M. genitalium* genome is so far the smallest discovered for any free-living organism, this mycoplasma has become the benchmark for approximation of the minimal gene set necessary for life. With that view of *M. genitalium* in mind, it is more interesting to examine the *M. genitalium* genes that have no orthologues in *U. urealyticum*. There are 74 *M. genitalium* proteins with some functional annotation that do not have orthologues in *U. urealyticum*. Ten of the *M. genitalium* genes missing in *U. urealyticum* are involved in energy metabolism. In terms of the

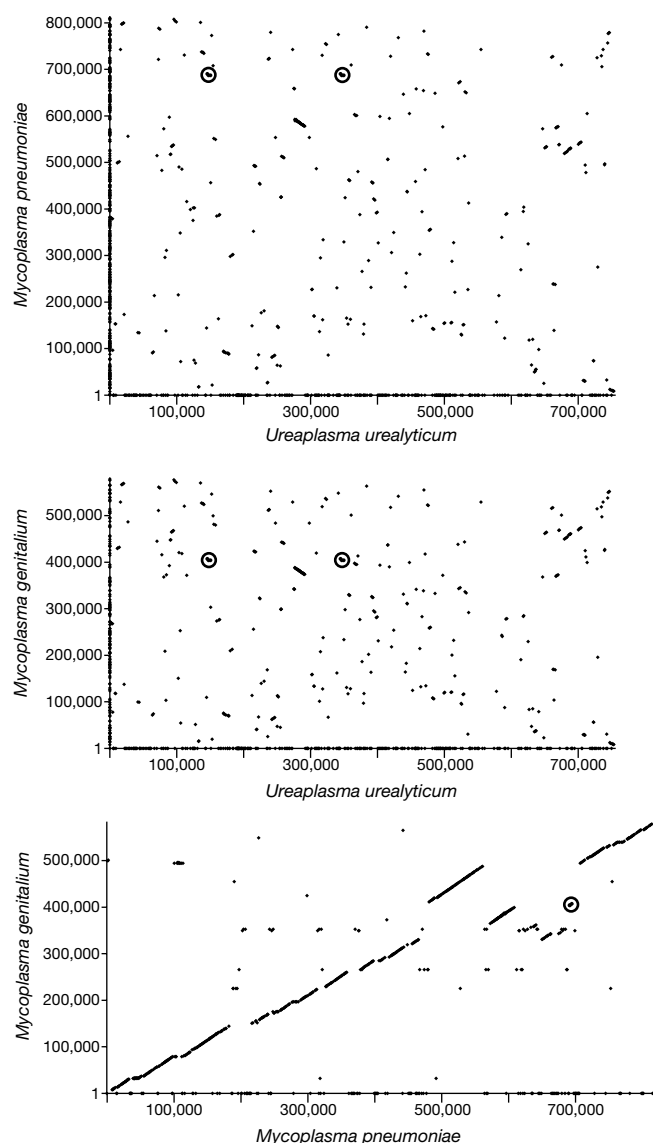


Figure 2 Conservation of gene order among *U. urealyticum*, *M. pneumoniae* and *M. genitalium*. Genes with corresponding orthologues in *U. urealyticum*, *M. genitalium* and *M. pneumoniae* were plotted with respect to their genome location. Base 1 for each organism corresponds to the starting base of the *dnaA* gene. The circled points represent the location(s) of the ribosomal RNA operon(s). There are two small syntenic regions between *U. urealyticum* and the other two organisms formed by the ribosomal proteins (275–291 kb) and the oligopeptide transporter genes (699–704 kb).

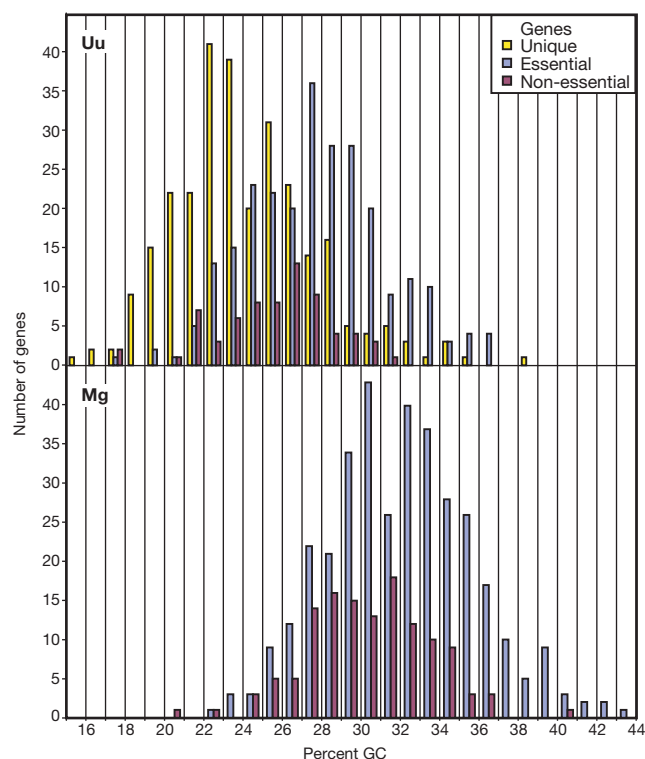


Figure 3 Distribution of *U. urealyticum* and *M. genitalium* genes according to percentage G+C. The base composition of protein-coding genes classified as unique, non-essential and potentially essential are plotted for *U. urealyticum* (Uu, top panel) and *M. genitalium* (Mg, bottom panel) (see text). The number of genes (indicated on the vertical axis) falling within a particular G+C frequency bin (indicated on the horizontal axis) is plotted for each of the gene classifications.

number of genes involved, ATP production by urea hydrolysis is a simpler process than is carbohydrate metabolism. The absence of other genes is more surprising. *U. urealyticum* lacks the heat shock protein/chaperonins GroEL and GroES. These double-ring prokaryotic chaperonins, which mediate protein folding within the cell, are found in all other sequenced microbial genomes, although they are not essential genes *in vitro*²⁶. Perhaps the most intriguing absence in ureaplasma is the cell-division protein FtsZ. All bacterial genomes sequenced to date, except for the chlamydias and the archaeon *Aeropyrum pernix* K1, contain *ftsZ*. Because chlamydia cell division takes place in eukaryotic cell vacuoles, FtsZ protein is not necessary. Among other eubacteria, FtsZ is presumed to be an essential component of the cell-division mechanism.

The mean base composition of the *U. urealyticum* genome is 25.5% G+C. The *U. urealyticum* genome is much more A+T-rich than any of the other prokaryotic genomes sequenced to date. Low G+C content is a general characteristic of the Mollicutes. The enzymatic cause of the mutation pressure leading to A+T enrichment in *U. urealyticum* is possibly a decreased capacity to remove uracil from DNA. Sources of uracil in DNA are mis-incorporation of dUTP by DNA polymerase and spontaneous deamination of deoxycytidine residues. Repair of the resulting G-U mismatches by DNA polymerase creates an A/T-biased mutation pressure. Two enzymes that could affect this process are dUTPase, which prevents dUTP from being incorporated into DNA, and uracil-DNA glycosylase, which removes uracil residues from DNA. We could not identify a dUTPase, nor could any dUTPase activity be detected in *U. urealyticum*¹⁶. *U. urealyticum* does have a uracil-DNA glycosylase; however, enzymatic analysis of ureaplasma extracts did not detect any uracil-DNA glycosylase activity²⁷. The uracil-DNA glycosylase of *U. urealyticum* and other low G+C mycoplasmas may remove uracil from DNA inefficiently. This is suggested by analysis of the uracil-DNA glycosylase of *Mycoplasma lactucae* (30% G+C), which has a K_m for uracil-containing DNA that is 40–1,000 times greater than values reported for other organisms²⁸.

Although the biased mutational pressure acts on the whole genome, the genome does not respond uniformly. Presumably, essential genes may have fewer sites where mutations are viable. We examined the G+C contents of sets of *U. urealyticum* genes that we believed were likely to be essential or non-essential to determine whether the high and low G+C contents were predictive of essentiality. We designated a subset of ureaplasma genes on the basis of an *M. genitalium* study that used transposon mutagenesis to classify genes as either non-essential or possibly essential. A set of 129 of the 480 protein-coding genes of *M. genitalium* was deleted using transposon mutagenesis, and thus shown to be non-essential. Of the remaining 351, at least 265 are probably essential²⁶. There are *U. urealyticum* homologues for 69 of the *M. genitalium* non-essential genes and for 255 of the possibly essential genes. That leaves 289 *U. urealyticum* genes that have no homologues in *M. genitalium* (referred to as 'unique' in this analysis). The three sets had different but overlapping G+C percentage distributions (Fig. 3). The histogram suggests that genes with lower G+C contents are likely to be expendable, at least under *in vitro* growth conditions. □

Methods

U. urealyticum serovar 3 isolate

We obtained the *U. urealyticum* serovar 3 isolate from E. A. Freundt (Institute of Medical Microbiology, University of Aarhus, Aarhus, Denmark). The sequenced isolate of *U. urealyticum* serovar 3 is available from the American Type Culture Collection as ATCC number 700970. We grew bacteria in 10B media supplemented with 20% fetal calf serum²⁹.

Genome sequencing

We used a hybrid sequencing strategy that we refer to as CROSS (for complete random and ordered shotgun sequencing). In the initial phase of CROSS, we generated a series of libraries containing *U. urealyticum* DNA that had been sheared or partially digested with restriction endonucleases and then size fractionated before insertion into vectors. We generated dye-primer sequences from lambda libraries containing 7–11-kilobase pair

(kbp) inserts, and plasmid libraries containing 1.5–2.5-kbp inserts. All sequencing was done using ABI 373 or 377 sequencers. When sequence coverage was at ~5×, we shifted from random shotgun to ordered shotgun sequencing²⁹. We identified those lambda clones that had poor sequence coverage between their two end sequences, and generated sequences from 1.5–2.5-kbp insert pUC18 libraries made from sheared polymerase chain reaction (PCR) amplicons of the lambda clone inserts. In this phase, we used a mixture of dye-primer and dye-terminator sequencing. Next, small mapped gaps (< 3.0 kbp) in the assembled sequence were closed by sequencing PCR products that covered those gaps. Unmapped gaps were mapped either by performing PCRs using all combinations of primers that annealed to the ends of all the contigs, or by sequencing directly from genomic DNA templates³⁰.

We used AutoAssembler 1.4 (PE Biosystems) for sequence assembly. Because AutoAssembler could not assemble all 11,291 sequences used in the project at once, we assembled the genome in ten segments by grouping those sequences from defined regions of the genome into subassemblies. Each DNA sequence was manually examined and edited. Every base determined from cloned *U. urealyticum* genomic DNA was sequenced using at least two different templates. Regions sequenced only from PCR products or genomic DNA templates were sequenced on both strands. We confirmed the sequence assembly by a combination of Southern blotting, PCR, and verifying that the map locations of the end sequences from lambda and plasmid inserts agreed with insert sizes. A description of the computational tools and databases that we used to annotate the individual genes can be found at <http://genome.microbio.uab.edu/uu> (see also Supplementary Information).

Received 13 March; accepted 26 July 2000.

- Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
- Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**, 4420–4429 (1996).
- Cassell, G. H., Waites, K. B., Watson, H. L., Crouse, D. T. & Harasawa, R. *Ureaplasma urealyticum* intrauterine infection: role in prematurity and disease in newborns. *Clin. Microbiol. Rev.* **6**, 69–87 (1993).
- Bradbury, J. M. International Committee on Systematic Bacteriology Subcommittee on Mollicutes: Minutes of the Interim Meeting 12 and 18 July 1996. Orlando, Florida, USA. *Int. J. System. Bacteriol.* **47**, 911–914 (1997).
- Smith, D. G., Russell, W. C., Ingledew, W. J. & Thirkell, D. Hydrolysis of urea by *Ureaplasma urealyticum* generates a transmembrane potential with resultant ATP synthesis. *J. Bacteriol.* **175**, 3253–3258 (1993).
- Shepard, M. C. Human mycoplasma infections. *Health Laboratory Science* **3**, 163–169 (1966).
- Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiol. Rev.* **59**, 451–480 (1985).
- Romano, N., La Licata, R. & Russo Alesi, D. Energy production in *Ureaplasma urealyticum*. *Pediatr. Infect. Dis.* **5**, 308–312 (1986).
- Neyrolles, O., Ferris, S., Behbahani, N., Montagnier, L. & Blanchard, A. Organization of *Ureaplasma urealyticum* urease gene cluster and expression in a suppressor strain of *Escherichia coli*. *J. Bacteriol.* **178**, 647–655 (1996).
- Smith, R. L., Thompson, L. J. & Maguire, M. E. Cloning and characterization of MgtE, a putative new class of Mg²⁺ transporter from *Bacillus firmus* OF4. *J. Bacteriol.* **177**, 1233–1238 (1995).
- Siewe, R. M. *et al.* Functional and genetic characterization of the (methyl)ammonium uptake carrier of *Corynebacterium glutamicum*. *J. Biol. Chem.* **271**, 5398–5403 (1996).
- Jayakumar, A., Epstein, W. & Barnes, E. M. J. Characterization of ammonium (methylammonium)/potassium antiport in *Escherichia coli*. *J. Biol. Chem.* **260**, 7528–7532 (1985).
- Soupe, E., He, L. D. Y. & Kustu, S. Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein. *Proc. Natl Acad. Sci. USA* **95**, 7030–7034 (1998).
- Bauminger, E. R. *et al.* Iron storage in *Mycoplasma capricolum*. *J. Bacteriol.* **141**, 378–381 (1980).
- Pollack, J. D., Merola, A. J., Platz, M. & Booth, R. L. J. Respiration-associated components of *Mollicutes*. *J. Bacteriol.* **146**, 907–913 (1981).
- Pollack, J. D., Williams, M. V. & McElhane, R. N. The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit. Rev. Microbiol.* **23**, 269–354 (1997).
- Ligon, J. V. & Kenny, G. E. Virulence of ureaplasma for mice. *Infect. Immun.* **59**, 1170–1171 (1991).
- Kilian, M., Brown, M. B., Brown, T. A., Freundt, E. A. & Cassell, G. H. Immunoglobulin A1 protease activity in strains of *Ureaplasma urealyticum*. *Acta Pathol. Microbiol. Scand. B* **92**, 61–64 (1984).
- De Silva, N. S. & Quinn, P. A. Localization of endogenous activity of phospholipases A and C in *Ureaplasma urealyticum*. *J. Clin. Microbiol.* **29**, 1498–1503 (1991).
- Zheng, X. *et al.* Small repeating units within the *Ureaplasma urealyticum* MB antigen gene encode serovar specificity and are associated with antigen size variation. *Infect. Immun.* **63**, 891–898 (1995).
- Upton, C. & Buckley, J. T. A new family of lipolytic enzymes? *Trends Biochem. Sci.* **20**, 178–179 (1995).
- Somerson, N. L., Walls, B. E. & Chanock, R. M. Hemolysin of *Mycoplasma pneumoniae*: tentative identification as a peroxidase. *Science* **150**, 226–228 (1965).
- Shepard, M. C. & Masover, M. C. in *The Mycoplasmas* (eds Barile, M. F. & Razin, S.) 451–494 (Academic, New York, 1979).
- Wren, B. W. *et al.* Characterization of a haemolysin from *Mycobacterium tuberculosis* with homology to a virulence factor of *Serpulina hyodysenteriae*. *Microbiology* **144**, 1205–1211 (1998).
- Himmelreich, R., Plagen, H., Hilbert, H., Reiner, B. & Herrmann, R. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* **25**, 701–712 (1997).
- Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **288**, 2165–2169 (1999).

27. Williams, M. V. & Pollack, J. D. in *DNA Replication and Mutagenesis* (eds Moses, R. C. & Summers, W. C.) 440–444 (American Society for Microbiology, Washington DC, 1988).
28. Williams, M. V. & Pollack, J. D. A mollicute (mycoplasma) DNA repair enzyme: purification and characterization of uracil-DNA glycosylase. *J. Bacteriol.* **172**, 2979–2985 (1990).
29. Chen, E. Y., Schlessinger, D. & Kere, J. Ordered shotgun sequencing, a strategy for integrated mapping and sequencing of YAC clones. *Genomics* **17**, 651–656 (1993).
30. Heiner, C. R., Hunkapiller, K. L., Chen, S. M., Glass, J. I. & Chen, E. Y. Sequencing multimegabase-template DNA with BigDye terminator chemistry. *Genome Res.* **8**, 557–561 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The authors thank D. Schlessinger for establishing the collaboration that led to this project, Y. Hale for growing the *U. urealyticum*; R. Belo, P. Babayan, K. Hunkapiller and T. Nguyen for assistance with DNA sequencing; and C.-N. Chen, S. Peterson, K. Ketchum and S. Payne for helpful discussions. This work was supported by PE Biosystems, the National Institute Allergy and Infectious Diseases of the National Institutes of Health, Eli Lilly and Company, and the University of Alabama at Birmingham, Department of Microbiology.

Correspondence and requests for materials should be addressed to J.I.G. (e-mail: Glass_John_I@Lilly.com). The annotated genome is available on (<http://genome.microbio.uab.edu/uu/uugen.htm>) and (<http://www.stdgen.lanl.gov/>). The sequence has been deposited with GenBank with accession number AF222894.

Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms

Pradeep K. Singh^{*†}, Amy L. Schaefer^{†‡}, Matthew R. Parsek^{†§}, Thomas O. Moninger^{||}, Michael J. Welsh^{*} & E. P. Greenberg^{†‡}

^{*} Howard Hughes Medical Institute & Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

[‡] Department of Microbiology, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

[§] Department of Civil Engineering, Northwestern University, Evanston, Illinois 60208, USA

^{||} Central Microscopy Research Facility, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

[†] These authors contributed equally to this work

The bacterium *Pseudomonas aeruginosa* permanently colonizes cystic fibrosis lungs despite aggressive antibiotic treatment^{1–3}. This suggests that *P. aeruginosa* might exist as biofilms—structured communities of bacteria encased in a self-produced polymeric matrix—in the cystic fibrosis lung^{1,4}. Consistent with this hypothesis, microscopy of cystic fibrosis sputum shows that *P. aeruginosa* are in biofilm-like structures. *P. aeruginosa* uses extracellular quorum-sensing signals (extracellular chemical signals that cue cell-density-dependent gene expression) to coordinate biofilm formation⁵. Here we found that cystic fibrosis sputum produces the two principal *P. aeruginosa* quorum-sensing signals; however, the relative abundance of these signals was opposite to that of the standard *P. aeruginosa* strain PAO1 in laboratory broth culture. When *P. aeruginosa* sputum isolates were grown in broth, some showed quorum-sensing signal ratios like those of the laboratory strain. When we grew these isolates and PAO1 in a laboratory biofilm model, the signal ratios were like those in cystic fibrosis sputum. Our data support the hypothesis that *P. aeruginosa* are in a biofilm in cystic fibrosis sputum. Moreover, quorum-sensing signal profiling of specific *P. aeruginosa* strains may serve as a biomarker in screens to identify agents that interfere with biofilm development.

Pseudomonas aeruginosa is the principal pathogen in the lungs of patients with cystic fibrosis (CF). Chronic colonization by this bacterium leads to progressive lung damage, and eventually respiratory failure and death in most CF patients^{2,3,6}. Once the CF lung has been colonized, even long-term antibiotic therapy does not eradicate *P. aeruginosa*^{2,3,7}. One hypothesis for the antibiotic resistance of *P. aeruginosa* in the CF lung is that this bacterium exists there as a biofilm, as bacterial biofilms are innately resistant to antimicrobial treatment¹.

Because little is known about the physiology of *P. aeruginosa* cells in biofilms, the hypothesis that *P. aeruginosa* in the CF lung is a biofilm has not been rigorously tested¹, and a marker for the biofilm-state has not been identified. Here, we first used transmission electron microscopy to examine CF sputum from patients infected with only *P. aeruginosa*. We prepared samples using a perfluorocarbon-based protocol to avoid the extraction of water-soluble material inherent in aqueous-based methods. We observed clusters of *P. aeruginosa* encased in a densely stained matrix (Fig. 1). Bacterial clusters were easily distinguished from the surrounding heterogeneous material containing host cells. We examined sputum samples from three different patients, and we did not observe bacteria outside these structures. This appearance is consistent with previous microscopic observations and consistent with the hypothesis that in CF sputum, *P. aeruginosa* exists in biofilms. As microscopic observations are open to interpretation, however, we sought physiological evidence of a biofilm lifestyle.

Biofilm development involves several steps. After attachment and multiplication on a surface, bacterial microcolonies form and then differentiate into characteristic antibiotic-resistant structures with cells encased in an extracellular polysaccharide matrix. In *P. aeruginosa*, differentiation is cued by one of two quorum-sensing signals that control the expression of dozens of *P. aeruginosa* genes in a cell-density-dependent fashion^{8,9}. The signals are *N*-(3-oxododecanoyl)-L-homoserine lactone (3OC12-HSL) and *N*-butyryl-L-HSL (C4-HSL). The signal required for microcolony differentiation in biofilms is 3OC12-HSL⁵. Although studies with mice indicate that quorum sensing is important for virulence of *P. aeruginosa*^{10,11},

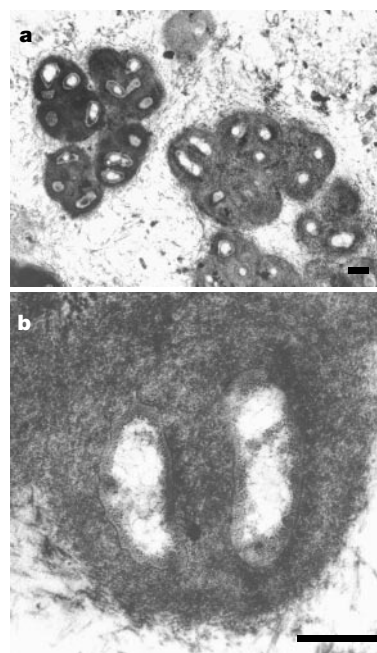


Figure 1 Transmission electron microscope images of *P. aeruginosa* in sputum. **a**, Low magnification; **b**, high magnification. *P. aeruginosa* are organized in clusters encased in a densely staining matrix. Scale bars, 1 μ m.

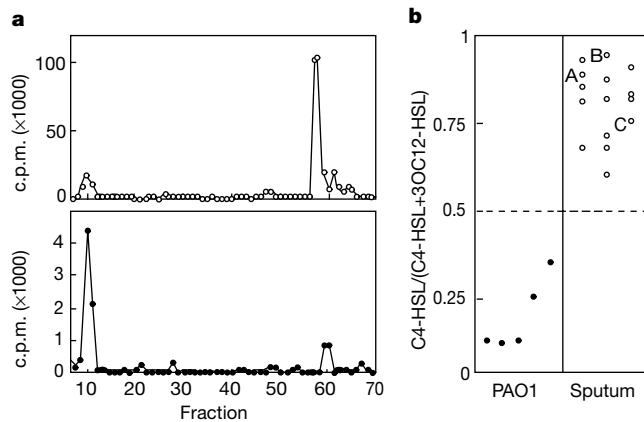


Figure 2 Acyl-HSL profiles of *P. aeruginosa* broth cultures and CF sputum samples. **a**, Examples of *P. aeruginosa* (top) and sputum (bottom) HPLC profiles. C4-HSL elutes in fractions 10–11; 3OC12-HSL elutes in fractions 57–61. The analysis of *P. aeruginosa* PAO1 was with a mid-logarithmic phase culture. **b**, C4-HSL as a fraction of the total C4-HSL plus 3OC12-HSL. From left to right, *P. aeruginosa* in early logarithmic ($A_{600} = 0.5$), mid-logarithmic ($A_{600} = 1$), late logarithmic ($A_{600} = 2$), stationary ($A_{600} = 4$) and late stationary phase ($A_{600} = 6$). Sputum samples were analysed as described in Methods. Three samples have been lettered because the predominant *P. aeruginosa* isolates from these samples have been subjected to a more extensive analysis.

it has not been shown that acyl-HSLs are produced by *P. aeruginosa* in lung secretions from CF patients or from any other human samples. This is due to limitations of the bioassays normally used to detect acyl-HSLs¹².

We have developed a radiometric technique that allows us to measure relative synthesis rates of acyl-HSLs in biological samples¹³. Using this technique, we showed that broth cultures of the common laboratory strain *P. aeruginosa* PAO1 produce 3OC12-HSL at a rate between 3 and 10 times that of C4-HSL production (Fig. 2). We then used the radiometric technique to test the hypothesis that *P. aeruginosa* in CF sputum produces acyl-HSL signals. In all cases, sputum samples from patients colonized with *P. aeruginosa* produced acyl-HSLs, but, in contrast to the broth-grown cultures of *P. aeruginosa* PAO1, the sputum samples produced more C4-HSL than 3OC12-HSL (Fig. 2). Clinical microbiology results showed that some sputum samples contained mucoid strains of *P. aeruginosa* that often colonize CF patients, others contained non-mucoid strains or more than one strain, and some sputum samples contained bacterial species in addition to *P. aeruginosa* (Table 1). We did not detect acyl-HSL synthesis in sputum from two CF patients free of *P. aeruginosa* but colonized by *Staphylococcus aureus*.

The differences in C4-HSL to 3OC12-HSL ratios in sputum

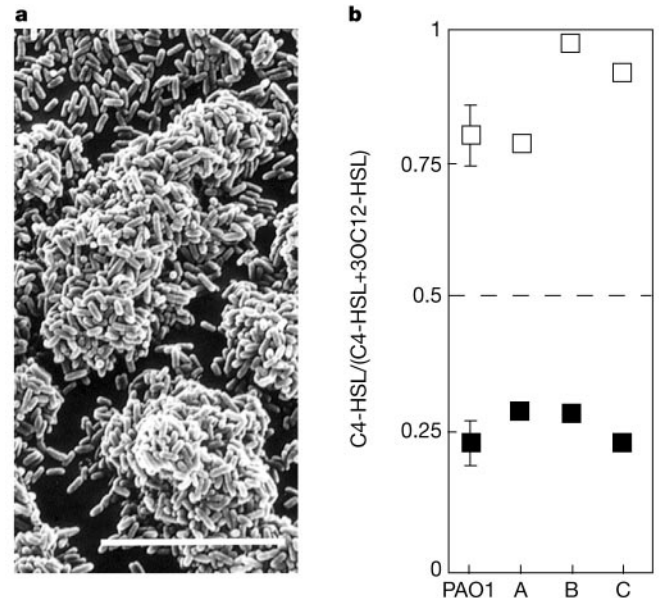


Figure 3 Comparisons of acyl-HSL ratios in biofilms with those in broth cultures.

a, Scanning electron micrograph of a *P. aeruginosa* PAO1 biofilm in a silicone tube. The tube was sliced open, and the sample was prepared for scanning electron microscopy. Microcolonies can be seen. Scale bar, 10 μm . **b**, C4-HSL as a fraction of the total C4-HSL plus C12-HSL in broth cultures (filled squares) and biofilms (open squares) of *P. aeruginosa* PAO1, and the most abundant bacterial type in sputum samples A, B and C (see Table 1). Data are mean $\pm$ s.d. for PAO1; broth and biofilm ratios were significantly different ($P < 0.001$). Ratios for CF isolates studied under different conditions were also statistically different ($P < 0.005$).

versus pure cultures grown in laboratory broth might relate to the different conditions in which *P. aeruginosa* is growing or it might be due to strain variation. To discriminate between these two possibilities, we analysed laboratory-grown broth cultures of nine clinical isolates from six of the sputum samples. Of the nine isolates, three produced more 3OC12-HSL than C4-HSL in culture broth (Fig. 3b). This finding indicates that for these isolates the conditions determined the ratio of acyl-HSLs. The other six isolates produced more C4-HSL than 3OC12-HSL when cultured in laboratory broth. In one isolate another acyl-HSL, C6-HSL, was the most abundant signal molecule. We also tested several non-CF clinical isolates of *P. aeruginosa*, and two *P. aeruginosa* environmental isolates, one from sewage and the other from a hot tub. All of these isolates produced more C4-HSL than 3OC12-HSL in laboratory broth (data not shown).

Two of the three CF clinical isolates for which the environmental

Table 1 Patient characteristics and bacterial species identified in sputum

Patient	Age (sex)	Years colonized	Antibiotic therapy*	Bacterial species	Colony-forming units per ml
A	29 (M)	6	–	Mucoid <i>P. aeruginosa</i>	10^8
				<i>Staphylococcus aureus</i>	10^7
B	27 (M)	>11	T, O	Mucoid <i>P. aeruginosa</i>	10^8
				<i>Stomatococcus mucilagines</i>	10^6
C	44 (M)	>13	O	Mucoid <i>P. aeruginosa</i>	10^8
				Nonmucoid <i>P. aeruginosa</i>	10^6
D	21 (M)	7	C, O	Mucoid <i>P. aeruginosa</i>	10^7
				<i>Xanthomonas maltophilia</i>	10^8
E	21 (F)	8	–	Mucoid <i>P. aeruginosa</i> type 1 and type 2	$10^7, 10^7$
F	18 (F)	8	O	Nonmucoid <i>P. aeruginosa</i> type 1 and type 2	$10^8, 10^8$
G	21 (M)	3	–	Mucoid <i>P. aeruginosa</i>	10^7
				<i>Haemophilus</i> sp.	10^6
H	32 (F)	10	T	Mucoid <i>P. aeruginosa</i> type 1 and type 2	$10^7, 10^7$
				Nonmucoid <i>P. aeruginosa</i>	10^6
I	57 (F)	49	T, O	Mucoid <i>P. aeruginosa</i> type 1 and type 2	$10^9, 10^8$

* T, inhaled tobramycin; C, ciprofloxacin; O, other antibiotic.

conditions determined whether the predominant acyl-HSL was C4-HSL or 3OC12-HSL came from sputum in which only one type of *P. aeruginosa* was isolated. The third isolate was the predominant type (the mucoid type, at 100-fold higher levels than the nonmucoid type) in the sputum from which it was isolated. As a test of the hypothesis that *P. aeruginosa* exists as a biofilm in CF lungs¹, we asked whether the strains that produced predominantly 3OC12-HSL in broth cultures would produce sputum-like acyl-HSL ratios when grown as biofilms. We established biofilms in silicone tubing (see Methods; and Fig. 3a), and measured C4-HSL:3OC12-HSL ratios (Fig. 3b). The acyl-HSL ratios produced by the biofilms showed the pattern characteristic of sputum. Thus, under the conditions of our experiments, the acyl-HSL profiles of these strains serve as a biomarker of their growth as a biofilm.

This analysis of the three biomarker strains isolated from CF sputum provides evidence that in CF sputum *P. aeruginosa* exists primarily as a biofilm. These strains produce C4-HSL in excess of 3OC12-HSL when grown in biofilms but not when grown in broth culture. The sputum from which they were isolated showed acyl-HSL profiles like the biofilms. This would not be expected if only a small portion of the bacteria in the sputum existed as a biofilm. Our findings are consistent with the hypothesis that *P. aeruginosa* in CF lungs exist as a biofilm. However, previous evidence in support of this hypothesis is scant^{1,14,15}, and the hypothesis has not addressed the question of whether a large or small percentage of the bacterial cells in a CF lung infection exists as a biofilm.

The radiometric technique allowed us to show that *P. aeruginosa* in CF sputum generates quorum-sensing signals, and allowed the identification of strains that show acyl-HSL profiles indicative of growth in a biofilm state. One might imagine that under different conditions either biofilms or routine laboratory cultures could exhibit different acyl-HSL profiles than the ones we observed. Nevertheless, with defined conditions the indicator strains should be useful in screening efforts aimed at identification of agents that block or disrupt normal biofilm activity. □

Methods

Sputum samples

Sputum (2–10 ml) was obtained from CF patients attending the University of Iowa CF Clinic for routine follow-up or with exacerbation of CF lung disease. Sputum was collected by voluntary expectoration. Patients gave informed consent and the protocol was approved by the University of Iowa Human Subjects Review Board. One millilitre of each sample was analysed at the University of Iowa Diagnostic Laboratory to determine the species and numbers of bacterial colony-forming units present.

Microscopic examination was performed on freshly expectorated sputum from three CF patients colonized only with *P. aeruginosa*. Osmium tetroxide (1%) was dissolved in perfluorocarbon (Fluorinert FC-72, 3M Corp.). Samples were gently immersed in the perfluorocarbon fixative for 1 h followed by three rinses in pure perfluorocarbon. Samples were then dehydrated in three washes of 100% ethanol, transitioned to Eponate-12 resin (Ted Pella Inc., CA) and cured overnight at 65 °C. Ultrathin sections were stained with uranyl acetate and lead citrate and imaged in a Hitachi H-7000 transmission electron microscope.

Analysis of acyl-HSL production

For analysis of acyl-HSL production in sputum, fresh samples were homogenized by vortexing with glass beads in an equal volume of PBS containing 2% N-acetylcysteine. The sputum was labelled with 10 µCi of ¹⁴C-carboxy-methionine (American Radiochemical Company) for 4 h with shaking at 37 °C. The acyl-HSLs were extracted with two equal volumes of acidified ethyl acetate (100 µl glacial acetic acid per litre solvent). The solvent was removed by evaporation under a gentle stream of N₂ gas, and the residue was suspended in 200 µl of 50% methanol in water and separated by methanol-gradient C₁₈ reverse phase high performance liquid chromatography (HPLC) (10–100% methanol) as described¹⁶. We determined radioactivity by scintillation counting. Corrections were made on the basis of the extraction coefficients of the acyl-HSLs. The acyl-HSL assignments were made by comparison with retention times of synthetic acyl-HSLs. The 4-h incubation period and the amount of radiolabel used were determined empirically to give sufficient product in our samples for ease of analysis. With shorter incubation periods (1 h) and less ¹⁴C-carboxy-methionine (5 µCi), there was more noise in the HPLC profiles but the C4-HSL:3OC12-HSL ratios were similar to those shown in Fig. 2.

For analysis of acyl-HSL production in broth cultures, we grew all strains of *P. aeruginosa* in Jensen's Medium with 0.3% glycerol at 37 °C with shaking¹⁷. Unless

otherwise indicated, we added 5 µCi of ¹⁴C-carboxy-methionine to cultures at an absorbance at 600 nm (*A*₆₀₀) of 1. After 10 min, we extracted and analysed the acyl-HSLs as described above.

Biofilms

We grew biofilms at 37 °C in silicone tubing (3.2 mm diameter) using a once-through continuous-flow system¹⁸. Jensen's Medium with 0.3% glycerol was pumped through a 19-cm section of tubing that had an internal volume of 1 ml at a rate of 0.2 ml min⁻¹. The tubing was inoculated with 1 ml of an overnight broth culture of *P. aeruginosa* by syringe, and after 15 min the flow of medium was initiated. To analyse acyl-HSL production, we halted flow and injected 10 µCi of ¹⁴C-carboxy-methionine in 1 ml of pre-warmed medium into the tube. After 30 min, we analysed the contents of the silicone tube as described above.

Samples were prepared for scanning electron microscopy by fixation in 2.5% glutaraldehyde followed by 1% osmium, dehydration in ethanol followed by hexamethyldisilazane, and air drying. The dried samples were mounted on aluminium stubs, sputter-coated with gold and palladium (60:40), and imaged with a Hitachi S-4000 electron microscope.

Differences in incubation time and in the amount of ¹⁴C-carboxy-methionine added to broth cultures and laboratory biofilms of *P. aeruginosa* PAO1 did not appear to influence C4-HSL:3OC12-HSL ratios. With both broth cultures and laboratory biofilms, we carried out experiments with 10 µCi of ¹⁴C-carboxy-methionine and with incubation times of up to 1 h; the acyl-HSL ratios were indistinguishable from those we report here. Furthermore, under a variety of growth conditions (different carbon sources and anaerobic growth) broth cultures of *P. aeruginosa* PAO1 produced 3OC12-HSL in excess of C4-HSL. Laboratory biofilms of *P. aeruginosa* PAO1 always produced more C4-HSL than 3OC12-HSL, regardless of incubation time (24–100 h) (M.R.P., A.L.S. and E.P.G., manuscript in preparation).

Received 17 July; accepted 8 August 2000.

- Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
- Burns, J. L., Ramsey, B. W. & Smith, A. L. Clinical manifestations and treatment of pulmonary infections in cystic fibrosis. *Adv. Pediatr. Infect. Dis.* **8**, 53–66 (1993).
- Hoiby, N. Antibiotic therapy for chronic infection of *Pseudomonas* in the lung. *Annu. Rev. Med.* **44**, 1–10 (1993).
- Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. Microbial biofilms. *Annu. Rev. Microbiol.* **49**, 711–745 (1995).
- Davies, D. G. *et al.* The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280**, 295–298 (1998).
- Welsh, M. J., Tsui, L.-C., Boat, T. F. & Beaudet, A. L. In *Metabolic and Molecular Basis of Inherited Diseases* Vol. III (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 3799–3876 (McGraw-Hill, New York, 1995).
- Mendelman, P. *et al.* Aminoglycoside penetration, inactivation, and efficacy in cystic fibrosis sputum. *Am. Rev. Respir. Dis.* **132**, 761–765 (1985).
- Whiteley, M., Lee, K. M. & Greenberg, E. P. Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **96**, 13904–13909 (1999).
- Pesci, E. C. & Iglewski, B. H. The chain of command in *Pseudomonas* quorum sensing. *Trends Microbiol.* **5**, 132–135 (1997).
- Tang, H. B. *et al.* Contribution of specific *Pseudomonas aeruginosa* virulence factors to pathogenesis of pneumonia in a neonatal mouse model of infection. *Infect. Immun.* **64**, 37–43 (1996).
- Rumrbaugh, K. P., Griswold, J. A. & Hamood, A. N. Contribution of the regulatory gene *lasR* to the pathogenesis of *Pseudomonas aeruginosa* infection of burned mice. *J. Burn Care Rehabil.* **20**, 42–49 (1999).
- Schaefer, A. L., Hanzelka, B. L., Parsek, M. R. & Greenberg, E. P. Detection, purification and structural elucidation of acylhomoserine lactone inducer of *Vibrio fischeri* luminescence and other related molecules. *Methods Enzymol.* **305**, 288–301 (2000).
- Schaefer, A. L., Greenberg, E. P. & Parsek, M. R. *Methods Enzymol.* (in the press).
- Costerton, J. W. The etiology and persistence of cryptic bacterial infections: a hypothesis. *Rev. Infect. Dis.* **6**, S608–S616 (1984).
- Costerton, J. W., Lam, J., Lam, K. & Chan, R. The role of the microcolony mode of growth in the pathogenesis of *Pseudomonas aeruginosa* infections. *Rev. Infect. Dis.* **5**, S867–S873 (1983).
- Pearson, J. P. *et al.* Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes. *Proc. Natl Acad. Sci. USA* **91**, 197–201 (1994).
- Jensen, J. E., Fecycz, T. & Campbell, J. N. Nutritional factors controlling exocellular protease production by *Pseudomonas aeruginosa*. *J. Bacteriol.* **144**, 844–847 (1980).
- Davies, D. G., Chakraborty, A. M. & Geesey, G. G. Exopolysaccharide production in biofilms: substratum activation of alginate gene expression by *Pseudomonas aeruginosa*. *Applied Environ. Microbiol.* **59**, 1181–1186 (1993).

Acknowledgements

We thank J. Launspach, C. Clark, and the University of Iowa Clinical Microbiology Laboratory for providing the data in Table 1. Funding was provided by the National Institute of General Medical Sciences, the National Heart, Lung, and Blood Institute, and the Cystic Fibrosis Foundation. A.L.S. was a NIH Graduate Trainee, M.R.P. was the recipient of an NIH Postdoctoral Research Service Award, and P.K.S. is the recipient of an NIH Mentored Physician Scientist Award and a Cystic Fibrosis Foundation Leroy Matthews Award.

Correspondence and requests for materials should be addressed to E.P.G. (e-mail: everett-greenberg@uiowa.edu).

Messenger RNA targeting of rice seed storage proteins to specific ER subdomains

Sang-Bong Choi*, Changlin Wang*, Douglas G. Muench†, Kenjiro Ozawa‡, Vincent R. Franceschi§, Yujia Wu* & Thomas W. Okita*

* Institute of Biological Chemistry, and §School of Biological Sciences, Washington State University, Pullman, Washington 99164, USA

† Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

‡ National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Japan 305-8602

Rice seeds, a rich reserve of starch and protein, are a major food source in many countries. Unlike the seeds of other plants, which typically accumulate one major type of storage protein, rice seeds use two major classes, prolamines and globulin-like glutelins. Both storage proteins are synthesized on the endoplasmic reticulum (ER) and translocated to the ER lumen, but are then sorted into separate intracellular compartments^{1,2}. Prolamines are retained in the ER lumen as protein bodies whereas glutelins are transported and stored in protein storage vacuoles. Mechanisms responsible for the retention of prolamines within the ER lumen and their assembly into intracisternal inclusion granules are unknown, but the involvement of RNA localization has been suggested³. Here we show that the storage protein RNAs are localized to distinct ER membranes and that prolamine RNAs are targeted to the prolamine protein bodies by a mechanism based on RNA signal(s), a process that also requires a translation initiation codon. Our results indicate that the ER may be composed of subdomains that specialize in the synthesis of proteins directed to different compartments of the plant endomembrane system.

Prolamine RNAs constitute the main storage protein RNA species on the surface of the prolamine protein bodies (PB-ER) in developing seeds^{4,5}, whereas the glutelin RNAs are the major species on the cisternal ER (C-ER)^{5,6}. The enrichment of prolamine RNAs on the PB-ER has been suggested to be an important step for the efficient folding and assembly of prolamine polypeptides into protein bodies³, a process involving the luminal chaperone BiP^{7,8}. Consistent with this view is that BiP is concentrated on the periphery of the prolamine protein bodies as compared to the rest of the ER^{7,8} (Fig. 1a).

To investigate if prolamine RNAs are specifically targeted to the PB-ER, *in situ* polymerase chain reaction after reverse transcription of RNA (RT-PCR) using prolamine-specific primers was performed on rice seed sections (Fig. 1b, left). Appropriate control experiments were performed to ensure the specificity of the *in situ* RT-PCR. Multiple ring-like structures of 1–2 µm in diameter were readily evident using confocal microscopy. These structures are identical to the structures derived by BiP immunofluorescence (Fig. 1a, right). In contrast, glutelin RNAs gave a dense, patchy staining pattern (Fig. 1b, right) clearly different from that displayed by prolamine RNAs.

To confirm that the fluorescent ring-like structures were prolamine protein bodies, we employed double labelling methods (Fig. 1c). Developing endosperm sections were subjected to *in situ* RT-PCR using either Alexa 594-dUTP (Fig. 1c, upper row) or Oregon Green 488-dUTP (Fig. 1c, lower row), and then stained with DiOC₆ (Fig. 1c, upper row) or rhodamine B hexyl ester (Fig. 1c, lower row). These apolar ER vital dyes stain the surface of the prolamine protein bodies much more intensely than the rest of the ER membrane complex, owing to their propensity to interact with

the hydrophobic prolamines⁹. In both instances, the distribution patterns of prolamine RNAs coincided with the surface of the prolamine protein bodies identified by vital staining (Fig. 1c, middle column). These results indicate that prolamine and glutelin RNAs are distributed on spatially distinct ER subdomains and that prolamine RNAs are localized specifically to the prolamine protein bodies.

To determine the nature of the signal determinant, that is, peptide versus RNA, responsible for the sorting of prolamine RNA to the PB-ER, we generated a series of transgenic rice plants expressing synthetic prolamine RNAs and investigated prolamine RNA localization by *in situ* reverse transcription (Fig. 2a–d). These synthetic prolamine genes—the wild type (Fig. 2a, pCAM36), signal peptide-deleted (Fig. 2b, pCAM37), substituted glutelin 3' untranslated region (UTR) (Fig. 2c, pCAM38) and translation-defective (AUG to UUG) (Fig. 2d, pCAM42)—were tagged with herpes simplex virus (HSV) or 6 × Histidine (His₆) epitope sequences between the coding sequence and 3' UTR junction to distinguish the synthetic messenger RNA from the native prolamine RNA. Transgenic plants were identified by northern blot analysis using antisense HSV or His₆ oligonucleotide probe (Fig. 2e). No RNA signals were detected from untransformed control plants. Immunoblot analysis of endosperm extracts of pCAM36 or pCAM38 transgenic plants using an HSV antibody showed a reactive polypeptide band at 14 kilodaltons (Fig. 2e), a molecular size identical to the native mature prolamine¹⁰. No visible polypeptide band was observed in pCAM42 plants expressing the translation-defective prolamine gene (data

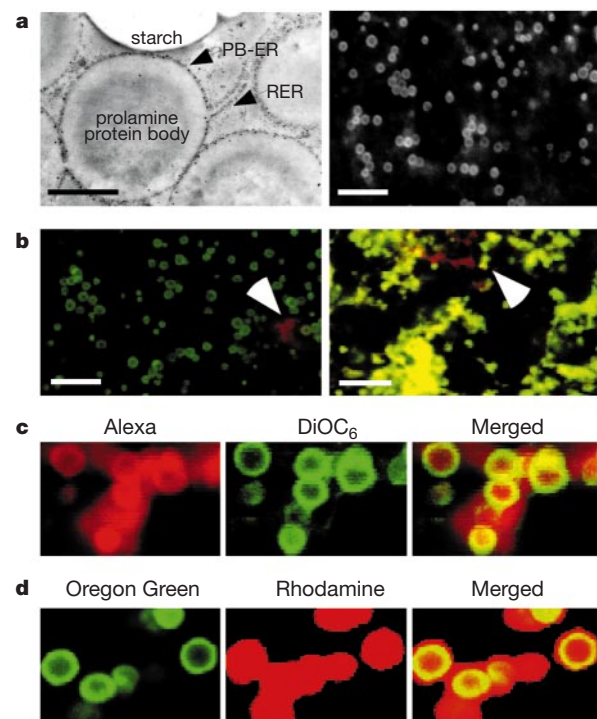


Figure 1 Localization of BiP, native prolamine and prolamine and glutelin transcripts in rice seed sections. **a**, Immunolocalization of BiP to the periphery of the prolamine protein bodies in thin- (left) and thick- (right) endosperm sections. BiP-antibody complexes were detected using (left) 15-nm gold particles conjugated to protein A, or (right) Alexa-594 conjugated to anti-rabbit IgG. Bar in left panel, 1 µm. Bar in right panel, 10 µm. RER, rough endoplasmic reticulum. **b**, Confocal laser scanning microscope images taken after *in situ* RT-PCR of seed sections in the presence of Oregon Green-dUTP with (left) prolamine- or (right) glutelin-specific primers. Seed sections were stained with propidium iodide to visualize chromatin (arrowhead). Bar, 10 µm. **c**, RT-PCR of seed sections with Alexa-dUTP and prolamine primers followed by DiOC₆ staining. **d**, Reaction with Oregon Green-dUTP and prolamine primers followed by rhodamine staining.

not shown). HSV-tagged prolamine protein encoded by pCAM37 was not detected with HSV antibody, presumably due to the mislocalization of the protein to the cytoplasm and pronounced instability in this location.

The His₆ or HSV epitope sequences may interfere with prolamine RNA localization; we therefore first investigated the distribution pattern of the synthetic wild-type prolamine transcripts in transgenic plants by *in situ* reverse transcription and confocal microscopy (Fig. 2a). Using an HSV primer, the synthetic wild-type prolamine RNA yielded a green fluorescence pattern similar to those observed for the endogenous prolamine transcripts (Fig. 1b and c, left). These fluorescently labelled ring structures were verified as prolamine protein bodies because they co-localized with the staining pattern obtained by rhodamine B hexyl ester (Fig. 2a, right). Under the same conditions, endosperm sections from untransformed plants showed only background levels of fluorescence (data not shown, but see Fig. 3a for a typical negative control). These results indicate that the addition of a small RNA segment between the coding and 3' UTR sequences has no effect on the localization of prolamine RNAs to the prolamine protein bodies.

The signal peptide has been suggested to be involved in the localization of prolamine RNAs to the prolamine protein bodies¹¹.

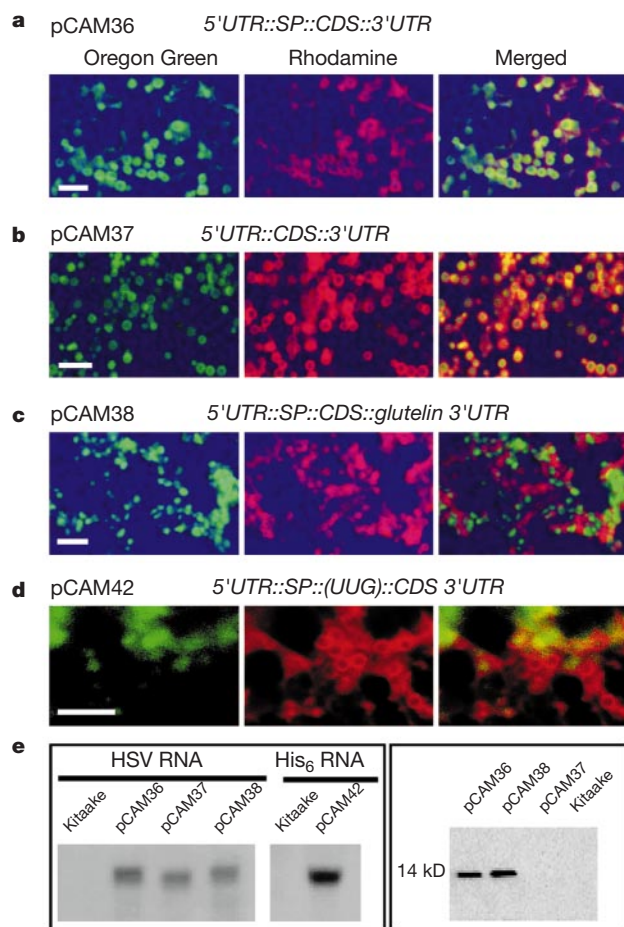


Figure 2 The distribution of synthetic prolamine RNAs in mid-developing rice endosperm sections. **a–d**, Seed sections from **(a)** pCAM36 (wild-type), **(b)** pCAM37 (signal peptide deleted), **(c)** pCAM38 (substituted glutelin 3' UTR), and **(d)** pCAM42 (AUG to UUG mutant) plants were subjected to *in situ* RT in the presence of Oregon Green-dUTP either with **(a–c)** HSV or **(d)** His₆ primer. The sections were subsequently treated with rhodamine B hexyl ester to stain the protein bodies. 5' UTR, prolamine 5' UTR; SP, prolamine signal peptide; SP(UUG), AUG to UUG mutant; CDS, coding sequence of prolamine; 3' UTR, prolamine 3' UTR. **e**, Northern blot analysis (left); immunodetection of HSV-tagged prolamine polypeptide (right). Kitsaake, untransformed rice. Bars, 10 μ m.

This proposal was based on the highly conserved sequence identity displayed between the signal peptides of the rice prolamines and maize zeins, two structurally distinct proteins that share a common storage site in the ER lumen. Confocal microscopic observations revealed, however, that the synthetic prolamine RNA lacking the signal peptide sequences (Fig. 2b) showed the same distribution pattern as the endogenous prolamine RNA (Fig. 1b, left), that is, they were localized to the prolamine protein bodies. This result indicates that the signal peptide and its corresponding RNA sequences are not essential for prolamine RNA accumulation.

Many localized RNAs have been shown to contain RNA zipcodes in the 3' UTR (see ref. 12 for review). To determine whether the prolamine 3' UTR serves a similar function, these sequences were replaced with the glutelin 3' UTR to generate pCAM38 (Fig. 2c). Analysis of the intracellular location of this transgene RNA showed a different distribution pattern from the wild type (Fig. 1b, left) and transgenic wild-type plants (Fig. 2a). Instead of ring structures observed in pCAM36 seed sections, pCAM38-expressing plants yielded fluorescent signals in small irregularly shaped patches (Fig. 2c, left), reminiscent of the distribution pattern obtained for glutelin RNAs (Fig. 1b, right). Based on these observations, we conclude that the prolamine 3' UTR contains RNA sequences essential for prolamine RNA localization to the prolamine protein bodies. Moreover, since pCAM38 plants synthesize a wild-type pre-prolamine polypeptide, prolamine RNA localization is independent of polypeptide sequences including the signal peptide.

Increasing evidence has been obtained that transporting RNA moves as a large complex containing many, if not all, of the components of the translational machinery (see ref. 13 for review). Indeed, a requirement for initiation of protein synthesis for accurate mRNA localization has been suggested¹⁴. To determine whether a translational process is involved in prolamine RNA transport, we investigated the distribution of prolamine transcripts lacking an AUG initiation codon in pCAM42 plants (Fig. 2d). The AUG-mutated RNAs were not distributed on the prolamine protein bodies, an observation that was consistently seen in several independent experiments with different individual transgenic plant lines. These results indicate that an initiation codon is required for faithful and accurate localization of prolamine RNAs to the

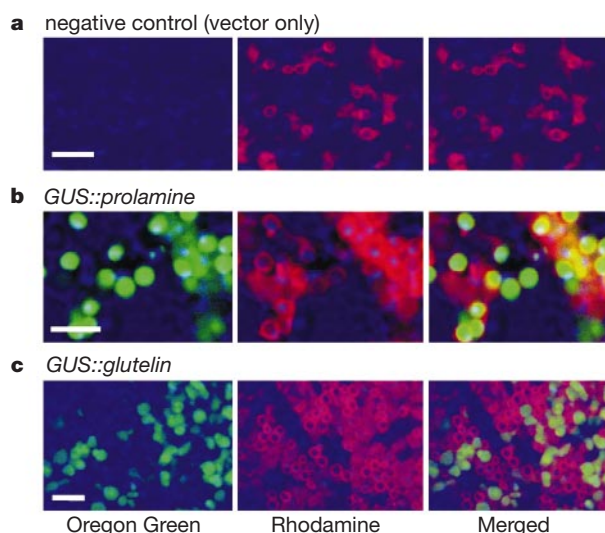


Figure 3 Visualization of chimaeric RNA transcripts in developing seeds by *in situ* RT-PCR. Images were obtained by the amplification of β -glucuronidase (GUS) transcripts using Oregon Green-dUTP and GUS-specific primers. **a**, Control plant transformed only with T-DNA vector. **b, c**, The GUS gene was fused upstream of **(b)** prolamine or **(c)** glutelin full-length cDNA, and expressed under the control of a rice prolamine promoter. Sections were stained with rhodamine B hexyl ester following PCR bars, 5 μ m.

prolamine protein bodies.

To demonstrate conclusively that storage protein RNA localization was dependent on RNA-based sorting mechanism, we analysed transgenic rice plants expressing a β -glucuronidase (GUS) reporter gene containing prolamine or glutelin RNA sequences located at the 3' UTR of the gene fusion (Fig. 3). GUS-specific primers were used to detect the hybrid RNA by *in situ* RT-PCR. No significant fluorescence signals were observed in control plants (Fig. 3a). While GUS transcripts are localized to the PB-ER in *GUS::prolamine* plants (Fig. 3b), such a distribution pattern was not evident in *GUS::glutelin* plants (Fig. 3c). These results confirm the RNA signal dependence of prolamine RNA localization to the PB-ER.

The exact signal domain responsible for localization of prolamine RNAs to the PB-ER has not been identified, but at least part of the sorting determinant resides in the 3' UTR. Although prolamine peptide sequences are not required, an AUG initiation codon is essential for RNA localization. This latter observation suggests that prolamine RNA moves as a ribonucleoprotein complex containing many components of the protein synthesis machinery especially factors required for translation initiation, that is, ribosome scanning of the initiation codon. Failure to form the translation initiation complex results in RNA mistargeting. As prolamine protein bodies are closely associated with the cortical cytoskeleton⁹, movement of the RNA-translation initiation complex is likely to be mediated by one or more cytoskeletal elements. The transport of ready-to-translate complexes would facilitate the efficient synthesis of large amounts of storage protein on the very restricted PB-ER, a process that probably expedites the formation of the prolamine granule in the ER lumen³. □

Methods

T-DNA vector construction and rice transformation

Synthetic prolamine genes were joined after PCR amplification of prolamine coding (CDS), 3' UTR and glutelin 3' UTR sequences from pProl14, prolamine 4a, and Gt2 genomic sequences respectively¹⁰. The HSV or His₆ sequence was added at the end of CDS during PCR. The DNA fragments were inserted downstream of the Gt1 promoter, and the promoter-insert fusion was cloned into the T-DNA transformation vector, pCAM-BIA1300. Rice (*Oryza sativa* cv. Kitaake) transformation was performed essentially as described¹⁵. Transgenic plants were grown in a controlled environment with an 11-hour, 26 °C day and a 13-hour, 22 °C night.

RNA and protein gel blot

Three micrograms of poly(A)⁺ RNA was resolved on a 1% formaldehyde agarose gel, transferred to a nylon membrane, and hybridized with ³²P end-labelled antisense HSV and His₆ oligonucleotides. Immunoblot of seed protein extracts were performed with either HSV monoclonal (Novagen) or His₆ (Invitrogen) antibody and enhanced chemiluminescence substrates (Pierce).

In situ RT-PCR and reverse transcription

Cryosections were prepared from 10–14 day old rice seeds, and fixed on glass slides for 30 min in 2.5% paraformaldehyde, 1.5% glutaraldehyde in 50 mM PIPES buffer, pH 7.2. The sections were then washed with phosphate-buffered saline (PBS) for 15 min followed by water rinses. Fixed sections were overlaid with RT-PCR reaction mixture containing 50 mM Tris-HCl, pH 9.0, 20 mM ammonium sulphate, 2.5 mM MgCl₂, 0.2 mM dATP, dCTP and dGTP, 10 μ M dTTP, 10 μ M Oregon Green 488-dUTP or Alexa 594-dUTP (Molecular Probes), 500 U ml⁻¹ MuLV reverse transcriptase (Stratagene), 50 U ml⁻¹ *Thermus thermophilus* (Tth) DNA polymerase (Epicentre), and 0.5 μ M of each primer (5'-CAGCTC GAGCAGCAATGAAG ATCATTTTCGT-3' and 5'-GTCTGAGCCAAGACACCGCCAAGGGTGGTA-3' for prolamine; 5'-GACAAGCTTGTCTGGCTTGGGA-TAAAGAATAAC-3' and 5'-CTGGAATCTAAACAATACTACTTAAAGGGGT-3' for glutelin; and 5'-CAGCGAAG AGGCAGTCAACGGGGAA-3' and 5'-CATTTGTTGCCCT-CCCTGTGCGGTT-3' for GUS). The RT-PCR was allowed to proceed at room temperature for 10 min followed by 60 °C for 20 min. The seed section was then subjected to 8–10 PCR cycles of 95 °C for 1 min, 72 °C for 1 min and 60 °C for 1 min. *In situ* reverse transcription was performed for 3 min at 94 °C, 20 min at 60 °C, and 5 min at 72 °C in the reaction solution containing 50 mM Tris-HCl, pH 9.0, 20 mM ammonium sulphate, 1.25 mM MgCl₂, 1 mM MnSO₄, PCR enhancer (Epicentre), 50 μ M each dATP, dGTP and dCTP, 10 μ M dTTP, 20 μ M Oregon Green 488-dUTP, 5 mM dithiothreitol, 150 U ml⁻¹ RNasin (5'-3', Inc.), 50 U ml⁻¹ Tth DNA polymerase and 0.1 μ M reverse primer (5'-GTCAAGCTTAATCCTCGGGGTCTTCCGGGG-3' for HSV-tagged plants; and 5'-TCAAGCTTAATGGTGATGGTGATGGTGAAGA-3' for His₆-tagged plants). Sections were stained with rhodamine B hexyl ester (final 0.1 μ M), DiOC₆ (0.5 μ M) or propidium

iodide (0.5 μ M), and then washed with 2 $\times$ SSC (0.3 M NaCl, 30 mM sodium citrate) for 15 min, 0.2 $\times$ SSC for 15 min twice, and then PBS for at least 10 hours at 4 °C.

Received 30 May; accepted 18 July 2000.

1. Krishnan, H. B., Franceschi, V. R. & Okita, T. W. Immunohistochemical studies on the role of the Golgi complex in protein body formation in rice seeds. *Planta* **169**, 471–480 (1986).
2. Yamagata, H. & Tanaka, K. The site of synthesis and accumulation of rice storage proteins. *Plant Cell Physiol.* **27**, 135–145 (1986).
3. Okita, T. W. & Rogers, J. C. Compartmentation of proteins in the endomembrane system of plant cells. *Annu. Rev. Plant Physiol. & Plant Mol. Biol.* **47**, 327–350 (1996).
4. Yamagata, H., Tamura, K., Tanaka, K. & Kasai, Z. Cell-free synthesis of rice prolamine. *Plant Cell Physiol.* **27**, 1419–1422 (1986).
5. Li, X., Franceschi, V. R. & Okita, T. W. Segregation of storage protein mRNAs on the rough endoplasmic reticulum membranes of rice endosperm cells. *Cell* **72**, 869–879 (1993).
6. Kim, W. T., Li, X. & Okita, T. W. Expression of storage protein multigene families in developing rice endosperm. *Plant Cell Physiol.* **34**, 595–603 (1993).
7. Li, X. *et al.* Rice prolamine protein body biosynthesis: A BiP-mediated process. *Science* **262**, 1054–1056 (1993).
8. Muench, D. G. *et al.* Molecular cloning, expression and subcellular localization of a BiP homolog from rice endosperm tissue. *Plant Cell Physiol.* **38**, 404–412 (1997).
9. Muench, D. G., Chuong, S. D. X., Franceschi, V. R. & Okita, T. W. Developing prolamine protein bodies are associated with the cortical cytoskeleton in rice endosperm cells. *Planta* **211**, 227–238 (2000).
10. Kim, W. T. & Okita, T. W. Structure, expression, and heterogeneity of the rice seed prolamines. *Plant Physiol.* **88**, 649–655 (1988).
11. Mitsukawa, N. & Tanaka, K. in *Rice Genetics II* (ed. Khush, G. S.) 503–512 (International Rice Research Institute, Manila, Philippines, 1991).
12. Bassell, G. J., Oleynikov, Y. & Singer, R. H. The travels of mRNAs through all cells large and small. *FASEB J.* **13**, 447–454 (1999).
13. Jansen, R. P. RNA-cytoskeletal associations. *FASEB J.* **13**, 455–466 (1999).
14. Dalglish, G. D., Veyrune, J. L., Accornero, N., Blanchard, J. M. & Hesketh, J. E. Localisation of a reporter transcript by the *c-myc* 3'-UTR is linked to translation. *Nucleic Acids Res.* **27**, 4363–4368 (1999).
15. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).

Acknowledgements

We thank R. S. Boston and F. Takaiwa for providing the maize BiP antibody and prolamine promoter clone. This work was supported in part by United States Department of Agriculture National Research Initiative Competitive Grant Program, by a National Science Foundation Grant, and by Washington State University College of Agriculture and Home Economics Hatch Project.

Correspondence and requests for materials should be addressed to T.W.O. (e-mail: tokita@wsu.edu).

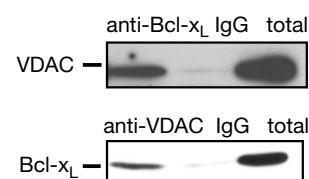
correction

Bcl-2 family proteins regulate the release of apoptogenic cytochrome *c* by the mitochondrial channel VDAC

Shigeomi Shimizu, Masashi Narita & Yoshihide Tsujimoto

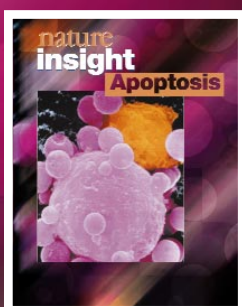
Nature **399**, 483–487 (1999).

Figure 1c of this paper was incorrect. The correct figure is reproduced below. □



nature insight

Apoptosis



Cover illustration

Coloured scanning electron micrograph showing a killer T lymphocyte (orange) inducing a cancer cell (mauve) to undergo programmed cell death or apoptosis. Mauve vesicles are apoptotic bodies emerging from the cancer cell. (Image courtesy of A. Liepins/SPL.)

Programmed cell death, or apoptosis, is currently one of the hottest areas of modern biology. It describes the orchestrated collapse of a cell, staging membrane blebbing, cell shrinkage, protein fragmentation, chromatin condensation and DNA degradation followed by rapid engulfment of corpses by neighbouring cells.

The excitement ensued when it became clear that apoptosis is an essential part of life for any multicellular organism and that the way in which most cells die is conserved from worm to mammal. Optimum body maintenance means that about 10 billion of our cells will die on a normal day just to counter the numbers of new cells that arise through mitosis. During development apoptosis helps to sculpture the body, shape the organs, and carve out fingers and toes. Both the nervous system and the immune system arise through overproduction of cells followed by the death of those that fail to establish functional synaptic connections or productive antigen specificities, respectively. Apoptosis is necessary to purge the body of pathogen-invaded cells, but is also needed to eliminate activated or auto-aggressive immune cells.

Such massacre has to be tightly regulated as too little or too much cell death may lead to pathology, including developmental defects, autoimmune diseases, neurodegeneration or cancer. Not surprisingly then that the hunt is on to understand which cells die when, why and how precisely, and to find drugs that interfere with specific steps along the pathway. Naturally, with over 50,000 publications on the subject to date (source: ISI-Web of Science), it is impossible to be comprehensive, but we hope that this *Nature Insight* provides our readers with a taster of the latest developments in this rapidly moving field.

Michael Hengartner sets the stage on page 770 and introduces the assassins and victims in this molecular 'murder mystery'. On page 777 Andrew Wyllie and co-workers discuss why and how DNA damage results in apoptosis in some cells but not in others and what the consequences are if cells with damaged genomes fail to die. Once a cell is committed to die, its corpse must be removed and destroyed by phagocytic cells, as discussed by John Savill and Valerie Fadok on page 784. Peter Krammer outlines the importance of apoptosis for the immune system on page 789, focusing on the role of the infamous CD95 death receptor. On page 796 Gerard Evan and colleagues outline the molecular mechanisms that bring about apoptosis during development of various organisms, and highlight the conservation of cell death mechanisms during evolution. The role of programmed cell death in the construction and pathological deconstruction of the brain is discussed on page 802 by Junying Yuan and Bruce Yankner. It is evident from reading these reviews that death or serious illness may result if cells that should die survive, or cells that should live die. There is clearly huge therapeutic potential, but will apoptosis deliver its promise to medicine? On page 810 Donald Nicholson discusses the opportunities and limitations of taking apoptosis from the bench to the clinic.

We are indebted to all of the contributors to this *Insight* for their considerable efforts, despite space and time constraints, in producing an enlightening and thought-provoking collection of reviews.

Marie-Thérèse Heemels Senior Editor

Ritu Dhand Insight Programme Editor
Liz Allen Publisher

The biochemistry of apoptosis

Michael O. Hengartner

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA (e-mail: hengartn@cshl.org)

Apoptosis — the regulated destruction of a cell — is a complicated process. The decision to die cannot be taken lightly, and the activity of many genes influence a cell's likelihood of activating its self-destruction programme. Once the decision is taken, proper execution of the apoptotic programme requires the coordinated activation and execution of multiple subprogrammes. Here I review the basic components of the death machinery, describe how they interact to regulate apoptosis in a coordinated manner, and discuss the main pathways that are used to activate cell death.

Multicellular animals often need to get rid of cells that are in excess, in the way, or potentially dangerous. To this end, they use an active dedicated molecular programme. As important as cell division and cell migration, regulated (or programmed) cell death allows the organism to tightly control cell numbers and tissue size, and to protect itself from rogue cells that threaten homeostasis.

Discovered and rediscovered several times by various developmental biologists and cytologists, programmed cell death acquired a number of names over the past two centuries¹. The term finally adopted is apoptosis, coined by Currie and colleagues in 1972 to describe a common type of programmed cell death that the authors repeatedly observed in various tissues and cell types². The authors noticed that these dying cells shared many morphological features, which were distinct from the features observed in cells undergoing pathological, necrotic cell death, and they suggested that these shared morphological features might be the result of an underlying common, conserved, endogenous cell death programme³.

Caspases: the central executioners

Most of the morphological changes that were observed by Kerr *et al.* are caused by a set of cysteine proteases that are activated specifically in apoptotic cells. These death proteases are homologous to each other, and are part of a large protein family known as the caspases⁴. Caspases are highly conserved through evolution, and can be found from humans all the way down to insects, nematodes and hydra^{5–7}. Over a dozen caspases have been identified in humans; about two-thirds of these have been suggested to function in apoptosis^{7,8}.

All known caspases possess an active-site cysteine, and cleave substrates at Asp-Xxx bonds (that is, after aspartic acid residues); a caspase's distinct substrate specificity is determined by the four residues amino-terminal to the cleavage site⁹. Caspases have been subdivided into subfamilies based on their substrate preference, extent of sequence identity and structural similarities.

Because they bring about most of the visible changes that characterize apoptotic cell death, caspases can be thought of as the central executioners of the apoptotic pathway. Indeed, eliminating caspase activity, either through mutation or the use small pharmacological inhibitors, will slow down or even prevent apoptosis⁷. Thus, blocking caspases can rescue condemned cells from their apoptotic fate — a fact that has not escaped the notice of the pharmaceutical industry (see review in this issue by Nicholson, pages 810–816).

It slices, it dices, and that's not all!

What exactly do the caspases do that is so important for apoptosis? Activation of caspases does not result in the wholesale degradation of cellular proteins. Rather, caspases selectively cleave a restricted set of target proteins, usually at one, or at most a few positions in the primary sequence (always after an aspartate residue). In most cases, caspase-mediated 'protein surgery' results in inactivation of the target protein (Box 1). But caspases can also activate proteins, either directly, by cleaving off a negative regulatory domain, or indirectly, by inactivating a regulatory subunit (Box 1).

Several important caspase substrates have been identified in recent years. One of the more exciting discoveries has been the elucidation of the mechanism of activation of the nuclease responsible for the famous nucleosomal ladder. First described by Wyllie¹⁰, this nuclease cuts the genomic DNA between nucleosomes, to generate DNA fragments with lengths corresponding to multiple integers of approximately 180 base pairs. The presence of this DNA ladder has been used (and abused) extensively as a marker for apoptotic cell death.

In an elegant series of experiments, the groups of Wang and Nagata showed that the DNA ladder nuclease (now known as caspase-activated DNase, or CAD) pre-exists in living cells as an inactive complex with an inhibitory subunit, dubbed ICAD (ref. 11). Activation of CAD occurs by means of caspase-3-mediated cleavage of the inhibitory subunit, resulting in the release and activation of the catalytic subunit^{12–14}.

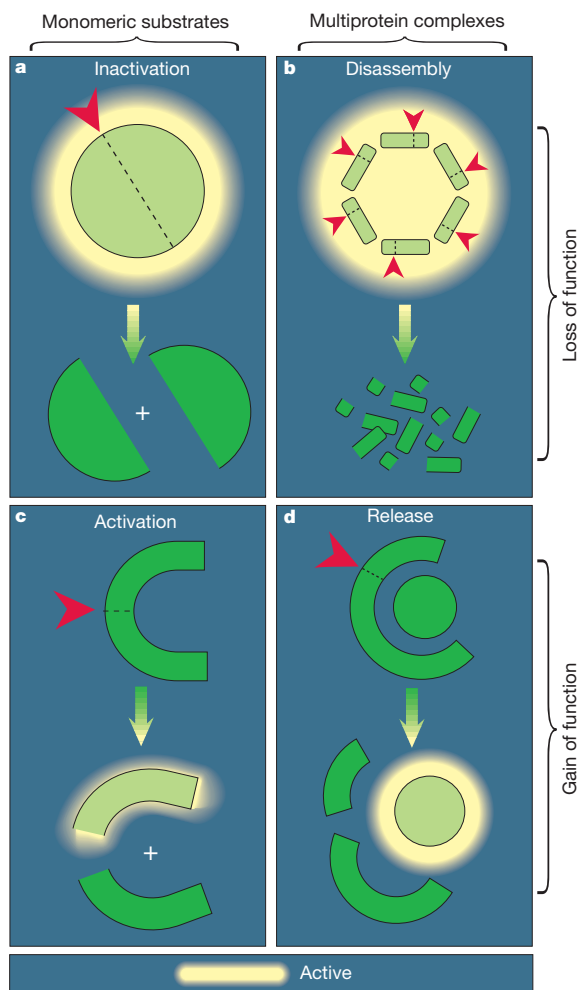
Caspase-mediated cleavage of specific substrates also explains several of the other characteristic features of apoptosis. For example, cleavage of the nuclear lamins is required for nuclear shrinking and budding^{15,16}. Loss of overall cell shape is probably caused by the cleavage of cytoskeletal proteins such as fodrin and gelsolin¹⁷. Finally, caspase-mediated cleavage of PAK2, a member of the p21-activated kinase family, seems to mediate the active blebbing observed in apoptotic cells. Interestingly, in this last case, caspase cleavage occurs between the negative regulatory subunit and the catalytic subunit, and results in a constitutive activation of PAK2 (ref. 18).

Close to 100 additional caspase substrates have been reported over the years, and there will certainly be many more^{7,19}. Why are there so many substrates? Perhaps apoptosis is just much more complicated than we currently believe. Indeed, several of the key apoptotic subprogrammes, such as cell shrinking and the emission of pro-engulfment signals (see review in this issue by Savill and Fadok, pages 784–788), are still poorly understood. Alternatively, it is possible that many of the described caspase substrates are not relevant substrates, but simply 'innocent bystanders' that get caught in the action.

Box 1

Camera, lights, action. Cut! Outcome of caspase activity

Proteolytic cleavage by caspases can lead to diverse results, depending on the nature of the substrate and the exact position of the cleavage site in the primary sequence. The simplest, and probably most frequent outcome is loss of biological activity (panels **a, b** in the figure below). Caspase substrates range from single polypeptide chain enzymes (for example, polyADP-ribose polymerase) to complex macromolecular structures (for example, the lamin network). Limited proteolysis by caspases can also result in a gain of biological activity (**c, d**). In some cases (for example, Bcl-2 or Bcl-x_L), the cleaved products antagonize the full-length protein (dominant-negative forms). In other cases, removal of inhibitory domains or subunits results in increased biological activity (for example, PAK2, Bid and CAD/ICAD).



According to this line of reasoning, there might be little selection against the presence of fortuitous caspase cleavage sites on irrelevant proteins, as the cell is about to stop functioning anyway. Further experimentation might allow this issue to be resolved.

How to activate a caspase

Given the great importance of caspases in the apoptotic process, it is reasonable to propose that a proper understanding of apoptosis will require us to understand how caspases are activated.

As is true of most proteases, caspases are synthesized as enzymatically inert zymogens. These zymogens are composed of three domains: an N-terminal prodomain, and the p20 and p10 domains,

which are found in the mature enzyme. In all cases examined so far, the mature enzyme is a heterotetramer containing two p20/p10 heterodimers and two active sites⁷. Although much has been made about the fact that active caspases are dimers containing two active sites, there is no obvious structural reason why this should be so, and it seems quite possible that caspases could exist as active monomers under the right conditions.

Three general mechanisms of caspase activation have been described so far. Each of them is described briefly below (see also Box 2).

Processing by an upstream caspase

Most caspases are activated by proteolytic cleavage of the zymogen between the p20 and p10 domains, and usually also between the prodomain and the p20 domain. Interestingly, all these cleavage sites occur at Asp-X sites — candidate caspase substrate sites — suggesting the possibility of autocatalytic activation⁹. Indeed, the simplest way to activate a procaspase is to expose it to another, previously activated caspase molecule (Box 2). This 'caspase cascade' strategy of caspase activation is used extensively by cells for the activation of the three short prodomain caspases, caspase-3, -6 and -7. These three downstream effector caspases are considered the workhorses of the caspase family, and are usually more abundant and active than their long prodomain cousins.

The caspase cascade is a useful method to amplify and integrate pro-apoptotic signals, but it cannot explain how the first, most upstream caspase gets activated. At least two other approaches are used to get the ball rolling.

Induced proximity

Caspase-8 is the key initiator caspase in the death-receptor pathway (see review in this issue by Krammer, pages 789–795). Upon ligand binding, death receptors such as CD95 (Apo-1/Fas) aggregate and form membrane-bound signalling complexes (Box 3). These complexes then recruit, through adapter proteins, several molecules of procaspase-8, resulting in a high local concentration of zymogen. The induced proximity model posits that under these crowded conditions, the low intrinsic protease activity of procaspase-8 (ref. 20) is sufficient to allow the various proenzyme molecules to mutually cleave and activate each other (Box 2). A similar mechanism of action has been proposed to mediate the activation of several other caspases, including caspase-2 and the nematode caspase CED-3 (ref. 21). Although forced crowding of zymogens clearly is sufficient in many cases to activate caspases²², it is a rather crude a way to control the fate of a cell. Whereas the basic concept is probably correct, additional levels of regulation surely must exist *in vivo* to modulate the process.

Association with a regulatory subunit

The most complex activation mechanism described so far is the one used by caspase-9. Unlike other caspases, proteolytic processing of procaspase-9 has only a minor effect of the enzyme's catalytic activity^{23,24}. Rather, the key requirement for caspase-9 activation is its association with a dedicated protein cofactor, Apaf-1 (Box 2).

Apaf-1 was identified through a biochemical approach as one of two proteins that are required for caspase-9 activation (the other being cytochrome c; see below)^{25,26}. Initially believed to be required only transiently, for caspase-9 activation, the Apaf-1/caspase-9 complex is now thought to actually represent the true active form of caspase-9 (ref. 23). Thus, we must view Apaf-1 not simply as a caspase-9 activator, but rather as an essential regulatory subunit of a caspase-9 holoenzyme. This holoenzyme — often referred to as the apoptosome — is a very large complex that might well contain several additional proteins^{27–29}.

In summary, effector caspases are usually activated proteolytically by an upstream caspase, whereas initiator caspases are activated through regulated protein–protein interactions. The actual molecular mechanisms mediating initiator caspase activation are still unclear and, most likely, much more complex than currently understood.

Regulated protein–protein interactions are in fact one of the underlying themes in apoptosis, and whole caspase activation pathways can be drawn without ever invoking a single enzyme (Box 3). I describe below some of the more commonly encountered interaction modules.

The handshakes that seal the fate

Each of the long-prodomain caspases contains in its prodomain a protein–protein interaction module, which allows it to bind to and associate with its upstream regulators. Caspase-8 and -10 contain a death-effector domain (DED), whereas caspase-2 and -9 contain a caspase activation and recruitment domain (CARD). These two domains share little sequence identity, but fold into very similar three-dimensional structures, consisting of six antiparallel α -helices arranged in a Greek key configuration³⁰. The same fold is also found in the death domain, a third protein interaction module present in several upstream regulators of apoptosis, such as CD95 and the adaptor molecular FADD (ref. 31). It seems likely that the death domain, DED and CARD are derived from a common ancestral domain³⁰.

The structure of the death domain, DED and CARD perfectly suits their function. The antiparallel helices bundle into a tight core, leaving exposed large surfaces onto which evolution has carved extended protein–protein interaction domains. The particular face of the module that is used for interaction varies greatly from one protein to the next^{31–33}. Work so far suggests that the death adaptor modules usually mediate intrafamily interactions (that is, death domain/death domain, DED/DED and CARD/CARD). However, structural analyses show that there is enough surface area left on death domains, DEDs and CARDS to also interact with other proteins. Indeed, death adaptor modules might well act as integration platforms, binding to several different proteins, which could modulate their dimerization and hence caspase activation.

Keep your friends close, but keep your enemies closer

Regulated protein–protein interactions are also key to the understanding of a second set of apoptotic regulators, the Bcl-2 family. This family has been divided into three groups, based on structural similarities and functional criteria (Box 4). Members of group I possess anti-apoptotic activity, whereas members of groups II and III promote cell death.

How do Bcl-2 family members control cell death? Bcl-2 family members seem to spend most of their time simply trying to block each other's next move. Many family members can homodimerize, but more importantly, pro- and anti-apoptotic members can form heterodimers^{34–36}. Because each Bcl-2 family member can interact with several other different members, large numbers of heterodimer combinations within a cell are possible. To a first approximation, heterodimerization can simply be thought as resulting in mutual neutralization of the bound pro- and anti-apoptotic proteins. Thus, the problem collapses into comparing overall levels of pro- and anti-apoptotic family members: cells with more pro-death proteins are sensitive to death; cells with an excess of protective family members are usually resistant.

But Bcl-2 proteins clearly will need to do more than just talk to each other if they are to influence cell death. What is the ultimate output from all these interactions? In the nematode *Caenorhabditis elegans*, the anti-apoptotic Bcl-2 homologue CED-9 protects cells from death by directly binding to and sequestering the Apaf-1 homologue CED-4 (ref. 37). Although this is an appealing scenario, a similar interaction has been very difficult, if not impossible, to detect in mammals, at least not under the conditions tested so far^{38–40}. Rather, the key function of Bcl-2 family members seems to be to regulate the release of pro-apoptotic factors, in particular cytochrome *c*, from the mitochondrial intermembrane compartment into the cytosol^{35,36}.

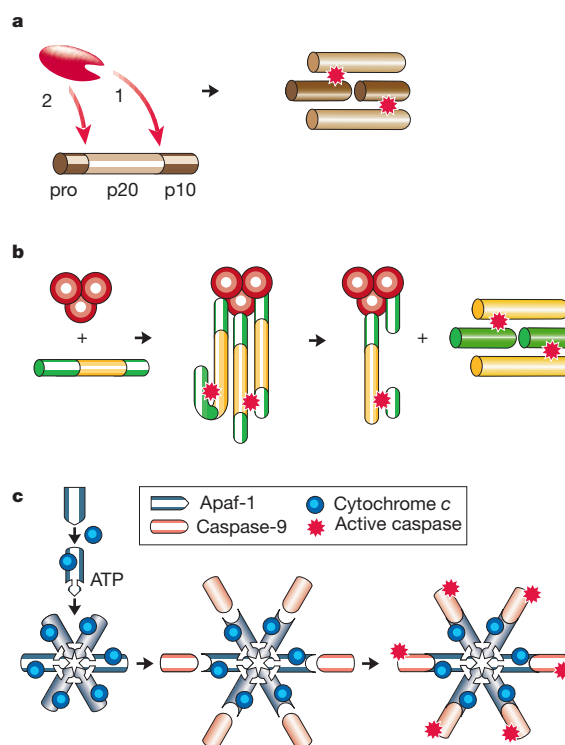
Mitochondria — the forum of death

The mitochondrion is not only the cell's powerhouse, it is also its arsenal. Mitochondria sequester a potent cocktail of pro-apoptotic proteins. Most prominent among these is cytochrome *c*, the humble electron carrier. Work over the past few years has revealed that cytochrome *c* is far from innocuous — in addition to its involvement in mitochondrial oxidative phosphorylation, the protein is one of the components (in addition to the adaptor protein Apaf-1) required for

Box 2

More than one way to skin a cat: mechanisms of caspase activation

Mechanisms of caspase activation include proteolytic cleavage by an upstream caspase (panel **a** in the figure below), induced proximity (**b**) and holoenzyme formation (**c**). Proteolytic cleavage by an upstream caspase is straightforward and effective, and is used mostly for activation of downstream, effector caspases. It is probably also used for induction of apoptosis by non-caspase proteases, such as granzyme B (see review in this issue by Krammer, pages 789–795). In the second mechanism, recruitment or aggregation of multiple procaspase-8 molecules into close proximity somehow results in cross-activation. The actual process is most probably more sophisticated and more tightly regulated than shown in panel **b**. In holoenzyme formation, cytochrome *c* and ATP-dependent oligomerization of Apaf-1 allows recruitment of procaspase-9 into the apoptosome complex. Activation of caspase-9 is mediated by means of conformational change, not proteolysis. Stoichiometry of the apoptosome is not known; it is shown in panel **c** as a hexamer solely for aesthetic reasons.



activation of caspase-9 in the cytosol²⁵.

Exactly how cytochrome *c* manages to cross the mitochondrial outer membrane is not yet known, but it is clear that the Bcl-2 family is intimately involved in the regulation of this process. For example, addition of pro-apoptotic Bcl-2 family members to isolated mitochondria is sufficient to induce cytochrome *c* release, whereas overexpression of Bcl-2 family members will prevent it^{35,36,41}.

How do Bcl-2 family members regulate cytochrome *c* exit? Several competing hypotheses have been advanced (Box 5); none of them has been proven definitively^{34–36,41}. The three basic models are as follows.

Bcl-2 members form channels that facilitate protein transport

Based on the structural similarity of Bcl-x_L to the pore-forming subunit of diphtheria toxin⁴², it has been suggested that Bcl-2 proteins might act by inserting, following a conformational change, into the outer mitochondrial membrane, where they could form channels or even large holes. Bcl-2 family members indeed can insert into

Box 3

The roads to ruin: two major apoptotic pathways in mammalian cells

The death-receptor pathway (left pathway in the figure opposite) is triggered by members of the death-receptor superfamily (such as CD95 and tumour necrosis factor receptor I). Binding of CD95 ligand to CD95 induces receptor clustering and formation of a death-inducing signalling complex. This complex recruits, via the adaptor molecule FADD (Fas-associated death domain protein), multiple procaspase-8 molecules, resulting in caspase-8 activation through induced proximity (see Box 2). Caspase-8 activation can be blocked by recruitment of the degenerate caspase homologue c-FLIP (ref. 61).

The mitochondrial pathway (right) is used extensively in response to extracellular cues and internal insults such as DNA damage (see review in this issue by Rich *et al.*, pages 777–783). These diverse response pathways converge on mitochondria, often through the activation of a pro-apoptotic member of the Bcl-2 family. Unlike Bcl-2,

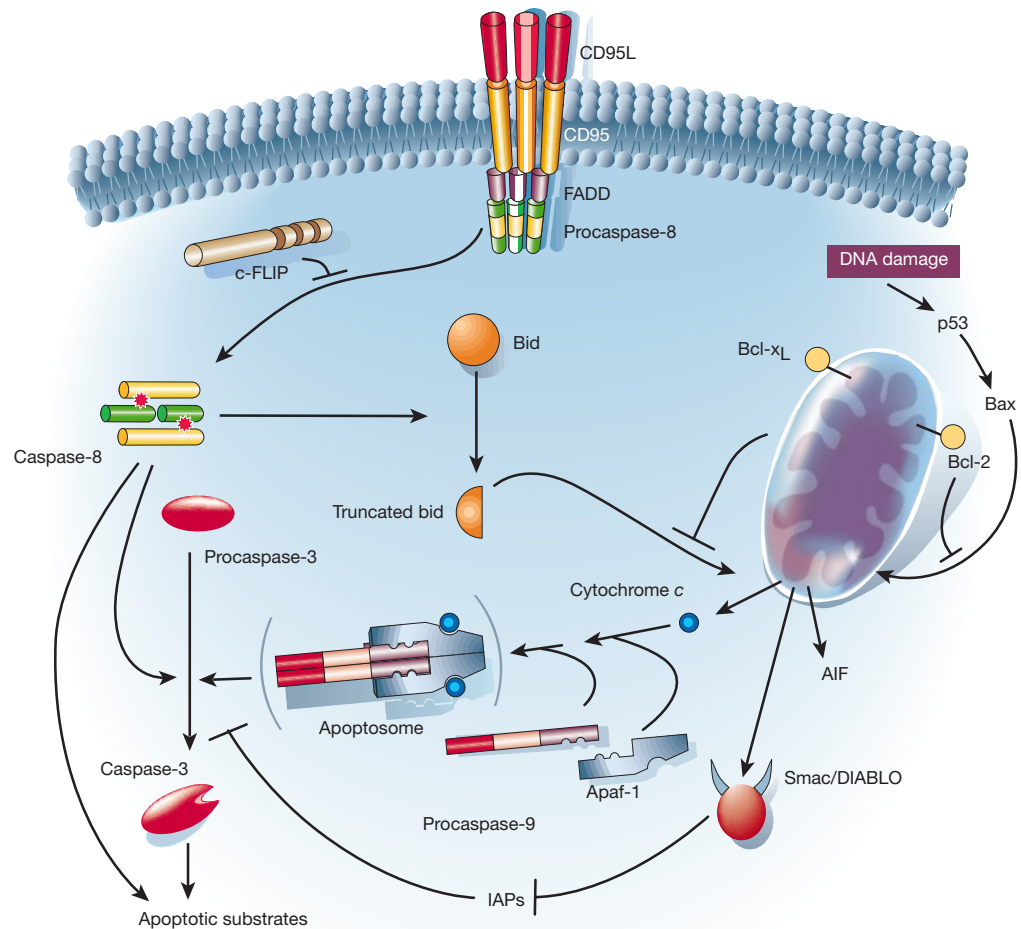
which seems to spend most if not all of its life attached to intracellular membranes, many group II and group III members, including Bax, Bad, Bim and Bid, can shuttle between the cytosol and organelles^{62–65}. The cytosolic forms represent pools of inactive, but battle-ready proteins. Pro-apoptotic signals redirect these proteins to the mitochondria, where the fight for the cell's fate will take place. Activation of pro-apoptotic members can occur through proteolysis, dephosphorylation and probably several other mechanisms^{35,36}.

Pro- and anti-apoptotic Bcl-2 family members meet at the surface of mitochondria, where they compete to regulate cytochrome *c* exit by a mechanism that is still debated (see text). If the pro-apoptotic camp wins, an array of molecules is released from the mitochondrial compartment. Principal among these is cytochrome *c*, which associates with Apaf-1 and then procaspase-9 (and possibly other proteins) to form the apoptosome. Heat-shock proteins act at multiple steps in the pathway to modulate apoptosis (not shown; see refs. 66, 67).

The death-receptor and mitochondrial pathways converge at the level of caspase-3 activation. Caspase-3 activation and activity is antagonized by the IAP proteins, which themselves are antagonized by the Smac/DIABLO protein released from mitochondria. Downstream of caspase-3, the apoptotic programme branches into a multitude of subprogrammes, the sum of which results in the ordered dismantling and removal of the cell.

Cross-talk and integration between the death-receptor and mitochondrial pathways is provided by Bid, a pro-apoptotic Bcl-2 family member. Caspase-8-mediated cleavage of Bid greatly increases its pro-death activity, and results in its translocation to mitochondria, where it promotes cytochrome *c* exit. Note that under most conditions, this cross-talk is minimal, and the two pathways operate largely independently of each other^{62,68}.

Clearly, additional death-inducing pathways must exist, as developmental apoptosis is by and large normal in mice defective in the caspase-8 and caspase-9 pathways^{7,52}.



synthetic lipid bilayers, oligomerize, and form channels with discrete conductances³⁴. But it unclear whether such channels would ever be big enough for proteins to pass through.

Bcl-2 members interact with other proteins to form channels

Bcl-2 family members interact with many proteins³⁴. One possibility is that pro-apoptotic family members recruit other mitochondrial outer membrane proteins into forming a large pore channel. A particularly attractive candidate for such a protein is the voltage-dependent anion channel (VDAC), as several Bcl-2 family members

can bind to it and regulate its channel activity⁴³. As the characterized pore size of the VDAC channel is too small to allow proteins to pass through, this model must assume that VDAC undergoes a significant conformational change upon binding to Bcl-2 family members.

Bcl-2 members induce rupture of the outer mitochondrial membrane

It is possible that the Bcl-2 family members control homeostasis of the mitochondria. In this model, apoptotic signals alter mitochondrial physiology (for example, ion exchange or oxidative phosphorylation) such that the organelle swells, resulting in the physical rupture

of the outer membrane and release of intermembrane proteins into the cytosol. The need to form a channel large enough for cytochrome *c* to pass through is thereby neatly bypassed as proteins can be assumed to simply diffuse through the tears in the lipid bilayer.

Mitochondrial homeostasis could be influenced directly by the Bcl-2 family members (for example, through the proposed intrinsic ion-channel activity mentioned above) or indirectly, through modulation of other mitochondrial proteins. The VDAC protein again is a prominent candidate for such regulation, as it is a subunit of the mitochondrial permeability transition pore (PTP), a large channel whose opening results in rapid loss of membrane potential and organellar swelling. Opening of the PTP quickly leads to cytochrome *c* release and apoptotic cell death, and pharmacological inhibitors of the PTP can act as potent inhibitors of cytochrome *c* release, and hence prevent apoptosis⁴⁴. However, cytochrome *c* exit can also occur in the absence of membrane potential loss^{41,44}, suggesting that the PTP cannot be the sole target of the Bcl-2 family proteins.

Cytochrome *c* exit is an almost universal feature of apoptotic cell death. However, in some cases, it is a very late event. For example, apoptosis induced by death receptors often bypasses the mitochondrial pathway⁴⁵. As might be expected, from the models discussed above, such deaths are relatively insensitive to protection by Bcl-2 (ref. 45), and cytochrome *c* release into the cytosol is likely to be the result of caspase activation, rather than its cause.

Cytochrome *c* is but one of a host of mitochondrial pro-death denizens. Also present in mitochondria and released upon induction of apoptosis are AIF (a flavoprotein with potent but relatively mysterious apoptotic activity⁴⁶), Smac/DIABLO^{47,48}, and several procaspases, including procaspase-2, -3 and -9 (ref. 44). Release of multiple death-promoting molecules might be necessary to insure swift and certain death — part of the plan to insure that activation of the apoptotic cascade is a one-way proposition.

Apoptotic antidotes and anti-antidotes

Is release of pro-death factors from mitochondria really the point of no return? Several lines of evidence suggest that cells can occasionally still be rescued at this stage — at least for a while. First, pharmacological inhibitors of caspases will often (but not always) rescue cells from apoptosis^{49,50}. Second, caspase-3- and caspase-9-knockout mice show reduced neuronal apoptosis during development and a significant defect in apoptosis following insult^{7,51,52}. Third, mammals (as well as the fruitfly *Drosophila* and some viruses) carry a family of genes that encode potent caspase inhibitors, known

as the inhibitors-of-apoptosis (IAP) proteins^{5,53}. There would be little reason for such proteins to exist if they could not influence the apoptotic process.

On the basis of the above, it might seem that cells suffer from a terminal case of indecisiveness when it comes to apoptotic cell death, letting apoptotic signalling go down endless trails but never fully committing (Box 3). But this impression would be wrong. Indeed, quite to the contrary, the apoptotic pathway contains a number of amplification steps and positive feedback loops that insure that a cell will either fully commit to death or completely abstain from it. For example, the fact that procaspases are caspase substrates insures rapid and complete conversion of a pool of proenzymes even if only a few molecules were initially activated⁸. Similarly, there is likely to be positive feedback between caspase activation and cytochrome *c* exit from mitochondria^{54,55}.

But positive feedback loops do require the presence of buffers and/or dampeners, or even the smallest perturbation would eventually lead to full activation and apoptotic death of the cell. The IAP proteins might well act as such dampeners. It is possible, for example, that IAPs are not meant to protect cells from frontal suicide assaults, but rather to squelch spurious spontaneous caspase activation.

This idea is further supported by the recent identification of a mammalian IAP inhibitor, known as Smac⁴⁷ (for second mitochondria-derived activator of caspases) or DIABLO⁴⁸ (for direct IAP-binding protein with low pI). As is the case for Reaper, Hid and Grim in *Drosophila* (ref. 56, and see review in this issue by Meier *et al.*, pages 796–801), Smac/DIABLO binds to IAP family members and neutralizes their anti-apoptotic activity. Most interestingly, Smac/DIABLO is normally a mitochondrial protein, but it is released into the cytosol in cells induced to die, presumably following the same exit route as cytochrome *c*.

Thus, if a cell is committed to apoptotic death such that it releases its mitochondrial contents, then Smac/DIABLO will sequester the IAP proteins and insure that they do not attempt to stop the programme in its tracks. By analogy, anti-apoptotic Bcl-2 family members can be thought of as buffers that minimize accidental release of mitochondrial contents. Several other buffer zones probably exist, waiting to be discovered.

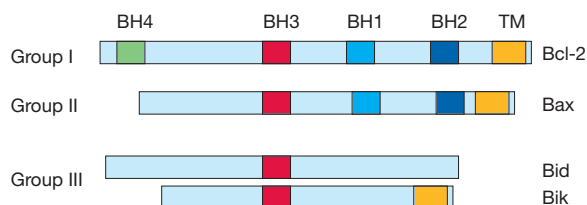
Would apoptosis, by any other name, be as sweet a death?

Is caspase activation the defining feature of apoptotic cell death? As I mentioned at the beginning of this review, most if not all of the morphological features used to initially describe apoptotic cell

Box 4

Gatekeepers and gatecrashers: Bcl-2 family members

Named after the founding member of the family, which was isolated as a gene involved in B-cell lymphoma (hence the name *bcl*; ref. 69), the Bcl-2 family is comprised of well over a dozen proteins, which have been classified into three functional groups^{35,36}. Members of the first group, such as Bcl-2 and Bcl-x_L, are characterized by four short, conserved Bcl-2 homology (BH) domains (BH1–BH4). They also possess a C-terminal hydrophobic tail, which localizes the proteins to the outer surface of mitochondria — and occasionally of the endoplasmic reticulum — with the bulk of the protein facing the cytosol. The key feature of group I members is that they all possess anti-apoptotic activity, and protect cells from death. In contrast, group II consists of Bcl-2 family members with pro-apoptotic activity. Members of this group, which includes Bax and Bak, have a similar overall structure to group I proteins, containing the hydrophobic tail and all but the most N-terminal, BH4 domain^{35,36}. Structure/function studies suggest that anti- versus pro-apoptotic activity is determined by relatively large regions of the protein, including two large α -helices that have been proposed to participate in membrane insertion (see text). Group III consists of a large and diverse collection of proteins whose only common feature is the presence of the ~12–16-amino-acid BH3 domain³⁵. Although some members of group III, including Bid, are indeed divergent homologues of Bcl-2 and Bax (refs 70, 71), others share little sequence or structural similarity with group I and II, suggesting that the BH3 domain in these proteins has arisen through convergent evolution⁴¹. Classification of such proteins as Bcl-2 family members is thus more a matter of convenience than a statement of presumed evolutionary relationship.

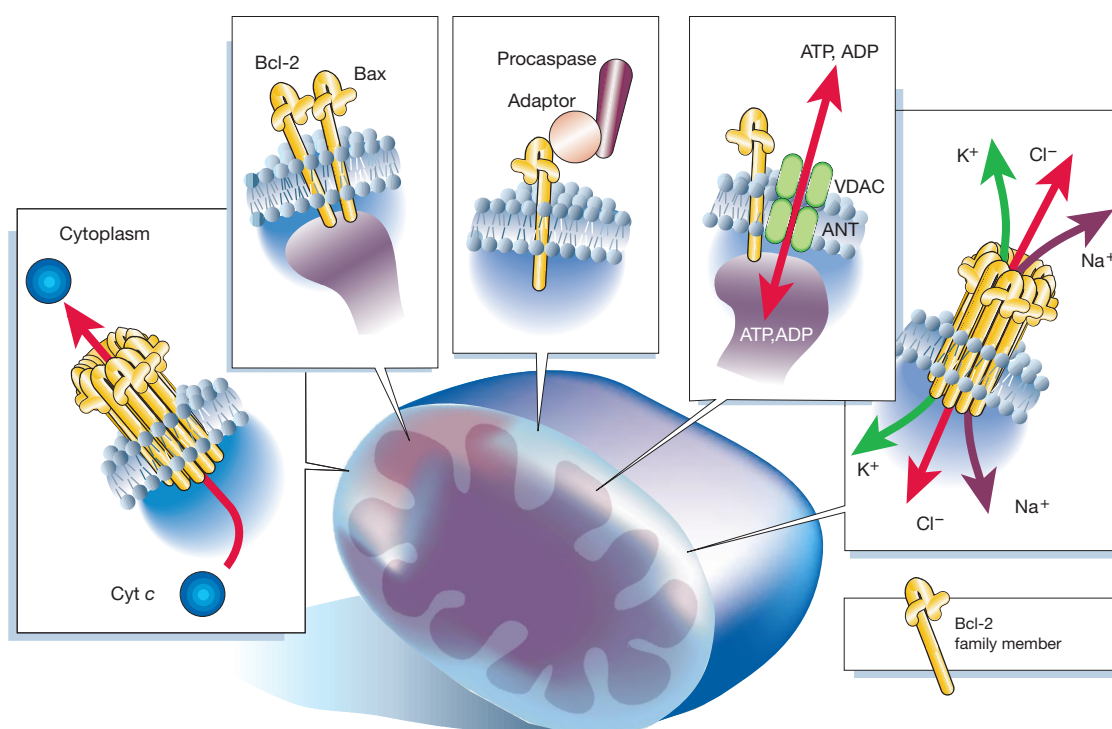


Box 5

Getting out the vote: possible mechanisms of action of Bcl-2 family members

Bcl-2 family members have been suggested to act through many different mechanisms^{34–36,41}. From left to right in the figure below, these include:

- Formation of a pore, through which cytochrome c (Cyt c) and other intermembrane proteins can escape.
- Heterodimerization between pro- and anti-apoptotic family members. Dimerization is achieved when the BH3 domain of one molecule binds into a hydrophobic pocket formed by the BH1, BH2 and BH3 domains of another family member⁷². Because of structural constraints, both homodimers and heterodimers are asymmetric molecules.
- Direct regulation of caspases via adaptor molecules, as has been described in *C. elegans*. Although the CED-4 homologue Apaf-1 is probably not a Bcl-2 family target, other adaptor proteins, such as BAR (ref. 73), the endoplasmic reticulum-localized protein Bap31 (ref. 74) and Aven (ref. 75), have been described in mammals.
- Interaction with other mitochondrial proteins, such as VDAC and the adenosine nucleotide transporter (ANT), either to generate a pore for cytochrome c exit, or to modulate mitochondrial homeostasis (for example, opening of the PTP).
- Oligomerization to form a weakly selective ion channel.



death are caspase-dependent. But the apoptotic programme is much more than just caspases, and in many cell types, activation of the apoptotic programme inevitably leads to death, with or without caspases⁵⁷. Programmed death of cells in which caspases have been blocked often bears little morphological similarity to apoptosis, and can even look surprisingly similar to classical necrotic cell death^{58,59}.

But that is not all. Physiological forms of cell death with non-apoptotic morphologies have been known for many years^{58,60}. How shall we classify such deaths? Atypical apoptosis? Necrosis? Non-apoptotic programmed cell death? Ideally, our final classification will be determined not by morphology, but by what molecular pathways are activated in the dying cell. This will require the development of ever more sophisticated assays for apoptotic proteins.

Conclusions

The field of apoptosis research took off as a result of the careful observations and astute deductions of a group of dedicated pathologists. As Yogi Berra said, "You can observe a lot by watching." Although many of the key apoptotic proteins have been identified, we still are mostly in the dark as to molecular mechanisms of action or activation of these proteins. Events downstream of caspases are murky, and there are kinds of cell death that have not even been touched yet.

There is much watching and much deducing left to do. Cell death will continue to be a lively field for the foreseeable future. □

1. Vaux, D. L. & Korsmeyer, S. J. Cell death in development. *Cell* **96**, 245–254 (1999).
2. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239–257 (1972).
3. Wyllie, A. H., Kerr, J. F. & Currie, A. R. Cell death: the significance of apoptosis. *Int. Rev. Cytol.* **68**, 251–306 (1980).
4. Alnemri, E. S. *et al.* Human ICE/CED-3 protease nomenclature. *Cell* **87**, 171 (1996).
5. Budihardjo, I., Oliver, H., Lutter, M., Luo, X. & Wang, X. Biochemical pathways of caspase activation during apoptosis. *Annu. Rev. Cell Dev. Biol.* **15**, 269–290 (1999).
6. Cikala, M., Wilm, B., Hobmayer, E., Bottger, A. & David, C. N. Identification of caspases and apoptosis in the simple metazoan Hydra. *Curr. Biol.* **9**, 959–962 (1999).
7. Earnshaw, W. C., Martins, L. M. & Kaufmann, S. H. Mammalian caspases: structure, activation, substrates, and functions during apoptosis. *Annu. Rev. Biochem.* **68**, 383–424 (1999).
8. Thornberry, N. A. & Lazebnik, Y. Caspases: enemies within. *Science* **281**, 1312–1316 (1998).
9. Thornberry, N. A. *et al.* A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J. Biol. Chem.* **272**, 17907–17911 (1997).
10. Wyllie, A. H. Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation. *Nature* **284**, 555–556 (1980).
11. Nagata, S. Apoptotic DNA fragmentation. *Exp. Cell Res.* **256**, 12–18 (2000).
12. Liu, X., Zou, H., Slaughter, C. & Wang, X. DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
13. Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).
14. Sakahira, H., Enari, M. & Nagata, S. Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis. *Nature* **391**, 96–99 (1998).
15. Rao, L., Perez, D. & White, E. Lamin proteolysis facilitates nuclear events during apoptosis. *J. Cell*

- Biol.* **135**, 1441–1455 (1996).
16. Buendia, B., Santa-Maria, A. & Courvalin, J. C. Caspase-dependent proteolysis of integral and peripheral proteins of nuclear membranes and nuclear pore complex proteins during apoptosis. *J. Cell Sci.* **112**, 1743–1753 (1999).
17. Kothakota, S. *et al.* Caspase-3-generated fragment of gelsolin: effector of morphological change in apoptosis. *Science* **278**, 294–298 (1997).
18. Rudel, T. & Bokoch, G. M. Membrane and morphological changes in apoptotic cells regulated by caspase-mediated activation of PAK2. *Science* **276**, 1571–1574 (1997).
19. Nicholson, D. W. Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death Differ.* **6**, 1028–1042 (1999).
20. Muzio, M., Stockwell, B. R., Stennicke, H. R., Salvesen, G. S. & Dixit, V. M. An induced proximity model for caspase-8 activation. *J. Biol. Chem.* **273**, 2926–2930 (1998).
21. Yang, X., Chang, H. Y. & Baltimore, D. Essential role of CED-4 oligomerization in CED-3 activation and apoptosis. *Science* **281**, 1355–1357 (1998).
22. Salvesen, G. S. & Dixit, V. M. Caspase activation: the induced-proximity model. *Proc. Natl Acad. Sci. USA* **96**, 10964–10967 (1999).
23. Rodriguez, J. & Lazebnik, Y. Caspase-9 and APAF-1 form an active holoenzyme. *Genes Dev.* **13**, 3179–3184 (1999).
24. Stennicke, H. R. *et al.* Caspase-9 can be activated without proteolytic processing. *J. Biol. Chem.* **274**, 8359–8362 (1999).
25. Li, P. *et al.* Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
26. Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
27. Beere, H. M. *et al.* Heat-shock protein 70 inhibits apoptosis by preventing recruitment of procaspase-9 to the Apaf-1 apoptosome. *Nature Cell Biol.* **2**, 469–475 (2000).
28. Cain, K. *et al.* Apaf-1 oligomerizes into biologically active approximately 700-kDa and inactive approximately 1.4-MDa apoptosome complexes. *J. Biol. Chem.* **275**, 6067–6070 (2000).
29. Cain, K., Brown, D. G., Langlais, C. & Cohen, G. M. Caspase activation involves the formation of the apoptosome, a large (approximately 700 kDa) caspase-activating complex. *J. Biol. Chem.* **274**, 22686–22692 (1999).
30. Hofmann, K. The modular nature of apoptotic signaling proteins. *Cell Mol. Life Sci.* **55**, 1113–1128 (1999).
31. Huang, B., Eberstadt, M., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature* **384**, 638–641 (1996).
32. Eberstadt, M. *et al.* NMR structure and mutagenesis of the FADD (Mort1) death-effector domain. *Nature* **392**, 941–945 (1998).
33. Zhou, P., Chou, J., Olea, R. S., Yuan, J. & Wagner, G. Solution structure of Apaf-1 CARD and its interaction with caspase-9 CARD: a structural basis for specific adaptor/caspase interaction. *Proc. Natl Acad. Sci. USA* **96**, 11265–11270 (1999).
34. Reed, J. C. Double identity for proteins of the Bcl-2 family. *Nature* **387**, 773–776 (1997).
35. Adams, J. M. & Cory, S. The Bcl-2 protein family: arbiters of cell survival. *Science* **281**, 1322–1326 (1998).
36. Antonsson, B. & Martinou, J. C. The Bcl-2 protein family. *Exp. Cell Res.* **256**, 50–57 (2000).
37. Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. Genetics of programmed cell death in *C. elegans*: past, present and future. *Trends Genet.* **14**, 410–416 (1998).
38. Pan, G., O'Rourke, K. & Dixit, V. M. Caspase-9, Bcl-XL, and Apaf-1 form a ternary complex. *J. Biol. Chem.* **273**, 5841–5845 (1998).
39. Hu, Y., Benedict, M. A., Wu, D., Inohara, N. & Nunez, G. Bcl-XL interacts with Apaf-1 and inhibits Apaf-1-dependent caspase-9 activation. *Proc. Natl Acad. Sci. USA* **95**, 4386–4391 (1998).
40. Hausmann, G. *et al.* Pro-apoptotic apoptosis protease-activating factor 1 (Apaf-1) has a cytoplasmic localization distinct from Bcl-2 or Bcl-x(L). *J. Cell Biol.* **149**, 623–634 (2000).
41. Gross, A., McDonnell, J. M. & Korsmeyer, S. J. BCL-2 family members and the mitochondria in apoptosis. *Genes Dev.* **13**, 1899–1911 (1999).
42. Muchmore, S. W. *et al.* X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **381**, 335–341 (1996).
43. Shimizu, S., Narita, M. & Tsujimoto, Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. *Nature* **399**, 483–487 (1999).
44. Loeffler, M. & Kroemer, G. The mitochondrion in cell death control: certainties and incognita. *Exp. Cell Res.* **256**, 19–26 (2000).
45. Scaffidi, C. *et al.* Two CD95 (APO-1/Fas) signaling pathways. *EMBO J.* **17**, 1675–1687 (1998).
46. Lorenzo, H. K., Susin, S. A., Penninger, J. & Kroemer, G. Apoptosis inducing factor (AIF): a phylogenetically old, caspase-independent effector of cell death. *Cell Death Differ.* **6**, 516–524 (1999).
47. Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
48. Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
49. Robertson, G. S., Crocker, S. J., Nicholson, D. W. & Schulz, J. B. Neuroprotection by the inhibition of apoptosis. *Brain Pathol.* **10**, 283–292 (2000).
50. Nicholson, D. W. ICE/CED3-like proteases as therapeutic targets for the control of inappropriate apoptosis. *Nature Biotechnol.* **14**, 297–301 (1996).
51. Zheng, T. S., Hunot, S., Kuida, K. & Flavell, R. A. Caspase knockouts: matters of life and death. *Cell Death Differ.* **6**, 1043–1053 (1999).
52. Wang, J. & Lenardo, M. J. Roles of caspases in apoptosis, development, and cytokine maturation revealed by homozygous gene deficiencies. *J. Cell Sci.* **113**, 753–757 (2000).
53. Miller, L. K. An exegesis of IAPs: salvation and surprises from BIR motifs. *Trends Cell Biol.* **9**, 323–328 (1999).
54. Green, D. & Kroemer, G. The central executioners of apoptosis: caspases or mitochondria? *Trends Cell Biol.* **8**, 267–271 (1998).
55. Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
56. Abrams, J. M. An emerging blueprint for apoptosis in *Drosophila*. *Trends Cell Biol.* **9**, 435–440 (1999).
57. Borner, C. & Monney, L. Apoptosis without caspases: an inefficient molecular guillotine? *Cell Death Differ.* **6**, 497–507 (1999).
58. Kitanaka, C. & Kuchino, Y. Caspase-independent programmed cell death with necrotic morphology. *Cell Death Differ.* **6**, 508–515 (1999).
59. Chautan, M., Chazal, G., Ceconi, F., Gruss, P. & Golstein, P. Interdigital cell death can occur through a necrotic and caspase-independent pathway. *Curr. Biol.* **9**, 967–970 (1999).
60. Depraetere, V. & Golstein, P. Dismantling in cell death: molecular mechanisms and relationship to caspase activation. *Scand. J. Immunol.* **47**, 523–531 (1998).
61. Irmeler, M. *et al.* Inhibition of death receptor signals by cellular FLIP. *Nature* **388**, 190–195 (1997).
62. Gross, A. *et al.* Caspase cleaved BID targets mitochondria and is required for cytochrome c release, while BCL-XL prevents this release but not tumor necrosis factor-R1/Fas death. *J. Biol. Chem.* **274**, 1156–1163 (1999).
63. Li, H., Zhu, H., Xu, C. J. & Yuan, J. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* **94**, 491–501 (1998).
64. Wolter, K. G. *et al.* Movement of Bax from the cytosol to mitochondria during apoptosis. *J. Cell Biol.* **139**, 1281–1292 (1997).
65. Puthalakath, H., Huang, D. C., O'Reilly, L. A., King, S. M. & Strasser, A. The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex. *Mol. Cell* **3**, 287–296 (1999).
66. Jaattela, M. Escaping cell death: survival proteins in cancer. *Exp. Cell Res.* **248**, 30–43 (1999).
67. Xanthoudakis, S. & Nicholson, D. W. Heat shock proteins as death determinants. *Nature Cell Biol.* **2**, E163–E165 (2000).
68. Yin, X. M. *et al.* Bid-deficient mice are resistant to Fas-induced hepatocellular apoptosis. *Nature* **400**, 886–891 (1999).
69. Tsujimoto, Y., Cossman, J., Jaffe, E. & Croce, C. M. Involvement of the bcl-2 gene in human follicular lymphoma. *Science* **228**, 1440–1443 (1985).
70. McDonnell, J. M., Fushman, D., Millman, C. L., Korsmeyer, S. J. & Cowburn, D. Solution structure of the proapoptotic molecule BID: a structural basis for apoptotic agonists and antagonists. *Cell* **96**, 625–634 (1999).
71. Chou, J. J., Li, H., Salvesen, G. S., Yuan, J. & Wagner, G. Solution structure of BID, an intracellular amplifier of apoptotic signaling. *Cell* **96**, 615–624 (1999).
72. Sattler, M. *et al.* Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **275**, 983–986 (1997).
73. Zhang, H. *et al.* BAR: an apoptosis regulator at the intersection of caspases and Bcl-2 family proteins. *Proc. Natl Acad. Sci. USA* **97**, 2597–2602 (2000).
74. Ng, F. W. *et al.* p28 Bap31, a Bcl-2/Bcl-XL- and procaspase-8-associated protein in the endoplasmic reticulum. *J. Cell Biol.* **139**, 327–338 (1997).
75. Chau, B. N., Cheng, E. H., Kerr, D. A. & Hardwick, J. M. Aven, a novel inhibitor of caspase activation, binds Bcl-xL and Apaf-1. *Mol. Cell* **6**, 31–40 (2000).

Acknowledgements

It is impossible to circumscribe the field of apoptosis in such a short review. I apologize to my many colleagues for having failed to cite their seminal papers and/or the critical results that clearly demonstrate their favourite model to be right. Many thanks to Y. Lazebnik for contributions to Box 3, and to Y.L., S. Lowe and members of the apoptosis community at Cold Spring Harbor Laboratory for many stimulating discussions. I dedicate this review to W. Hengartner on the occasion of his retirement from active mathematical duty.

Defying death after DNA damage

Tina Rich, Rachel L. Allen & Andrew H. Wyllie

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK (e-mail: ahw21@cam.ac.uk)

DNA damage frequently triggers death by apoptosis. The irreversible decision to die can be facilitated or forestalled through integration of a wide variety of stimuli from within and around the cell. Here we address some fundamental questions that arise from this model. Why should DNA damage initiate apoptosis in the first place? In damaged cells, what are the alternatives to death and why should they be selected in some circumstances but not others? What signals register DNA damage and how do they impinge on the effector pathways of apoptosis? Is there a suborganellar apoptosome complex effecting the integration of death signals within the nucleus, just as there is in the cytoplasm? And what are the consequences of failure to initiate apoptosis in response to DNA damage?

With few known exceptions, the terminal apoptotic programme of mammalian cells depends on the activation of intracellular caspases and their modification of protein substrates within the nucleus and cytoplasm (see review in this issue by Hengartner, pages 770–776). Two processes lie immediately upstream of these effector events. The first is the activation of the receptor-mediated death-signalling pathways that ultimately trigger caspase-8 and are exemplified by the interaction of CD95 (Apo-1/Fas) with its ligand (see ref. 1 and review by Krammer, pages 789–795). The second originates from mitochondria, which are central targets for intracellular oxidative stress. Stressed mitochondria release a set of molecules — cytochrome *c*, Apaf-1 and apoptosis-initiating factor — two of which contribute to a suborganellar molecular cluster (the apoptosome), which is then responsible for the activation of caspase-9 (ref. 2). This pathway can be profoundly influenced by both pro-apoptotic and anti-apoptotic members of the Bcl-2 family, which are in turn modified, in response to local survival factors, by phosphoinositide 3-kinase (PI(3)K) and Akt (ref. 3).

This article is concerned with the relation between DNA damage and the terminal apoptotic programme. Because its normal functions demand structural and sequence integrity over many hundreds of millions of non-redundant base pairs, the mammalian genome presents an enormous target to genotoxic agents. Moreover, DNA is highly reactive and is easily altered by cell processes such as oxidation. One estimate is that a mammalian genome undergoes about 100,000 modifications per day, each bearing a finite probability of residual damage⁴. The chromatin proteins in which DNA is embedded might afford some protection from this damage, and powerful repair mechanisms exist to restore DNA structure and sequence once damage has occurred. Nevertheless, the vital processes of replication, transcription and even repair itself require chromatin rearrangement, implying periods during which DNA vulnerability might be enhanced. Apoptosis is numerically important as one possible outcome of such DNA damage. Why is it necessary for cells to adopt this seemingly wasteful strategy alongside repair?

Why should DNA damage initiate apoptosis?

Cells differ hugely in their responses to DNA damage⁵. Whereas splenic lymphocytes in fetus and adult readily initiate apoptosis after exposure to ionizing radiation (which delivers double-strand DNA breaks to all cells),

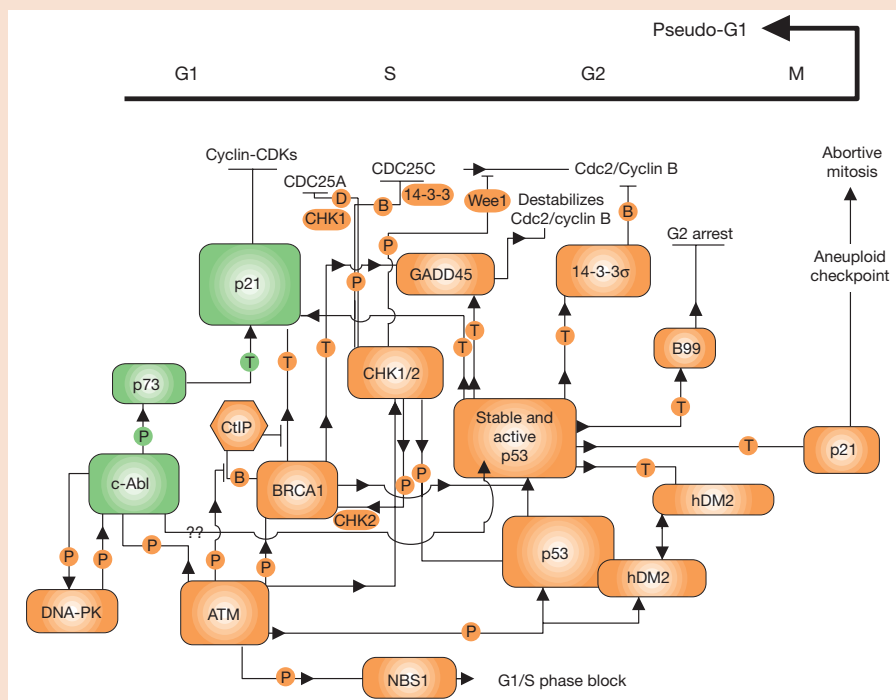
apoptosis forms no part of the response of cardiac myocytes to radiation at any stage of development. Mesenchymal cells of primordial cartilage are sensitive to radiation but become resistant on full differentiation. The post-replicative epithelial cells of the adult intestinal crypt are resistant to apoptosis in response to ionizing radiation and many other DNA-damaging agents, yet replicative cells of the same lineage, just a few hours earlier in their life history, and one cell position deeper in the crypt, are acutely sensitive to both radiation-induced and drug-induced apoptosis⁶. But in the thymic cortex, small CD4⁺/CD8⁺ lymphocytes have completed their last division yet remain sensitive to apoptosis after DNA damage and other stimuli^{7,8}. These examples emphasize that apoptosis is not an inevitable consequence of DNA damage. So why should they be coupled at all?

Although apoptosis is uniformly present in metazoans, both as a developmental programme (see review in this issue by Meier *et al.*, pages 796–801) and — in some circumstances — as an injury response, there is still controversy over its existence in unicellular organisms⁹. Certainly, the yeast genome does not encode a protein that, in metazoans, has the capacity to transduce DNA injury stimuli into the apoptotic programme with great efficiency: p53 (ref. 10). Even in mammals, p53 is frequently activated by DNA injury to serve other purposes than the initiation of apoptosis⁵. This raises the possibility that the coupling of DNA damage and apoptosis might be a strategy, adapted from other injury responses, to cope with certain problems of tissue organization.

Metazoan tissues depend absolutely on the ability of their constituent cells to relate to each other. Through cell–cell and cell–matrix communication, the functions of replication, differentiation and movement are orchestrated and topologically constrained. Some of these processes are difficult to reverse or rectify in the event of failure, yet failure is never far away. A half dosage of just one gene — APC, which encodes the oncosuppressor protein Adenomatous Polyposis Coli — renders the intestinal epithelium susceptible to the development of cells with inaccurate perceptions of polarity and position, and loss of restraint in replication: the founder cells of adenomas¹¹. It is possible that cells in metazoan tissues safeguard all the important phase transitions in their lifespans against injury-induced genetic error by linking them conditionally to a death programme already in use for pruning cellular genealogical trees and sculpting organs during development.

Modes of death that are less proactive than apoptosis are intolerably disruptive to tissue organization. Furthermore,

Figure 1 ATM, checkpoints and the cell cycle. DNA damaged by ionizing radiation can be sensed by ATM, triggering a cascade of downstream pathways to arrest the cell cycle. ATM-proximal events are phosphorylation reactions (denoted by P) that can lead to downstream transactivation events (T), degradations (D) or inhibitory blockades (B). p21 can block the G1/S transition and prevents aneuploidy. Multiple proteins transactivated by p53 block the S/G2 transition. Green boxes mark an auxiliary pathway that uses p73 to activate p21. S-phase blocking is also achieved by the phosphorylation of CDC25C by CHK, resulting in its cytoplasmic sequestration (by 14-3-3). Similarly, Cdc2 and cyclin B are inhibited by 14-3-3 σ .



the presence of free DNA ends in a cell that retains a capacity for DNA repair leads to the activation of poly(ADP-ribose) polymerase (PARP) and the consequent exhaustion of cellular energy supplies¹². The resulting clusters of dead cells would distort the critical ongoing cell–cell and cell–matrix signalling of a metazoan tissue. In contrast, apoptosis is designed to delete cells from tissues rapidly, tagging them for phagocytosis and recycling their constituent molecules, while neatly delaying energy exhaustion by uncoupling (through caspase activation) the catalytic and DNA-binding domains of PARP. By implication, the threshold for the activation of apoptosis in response to DNA damage can be set low: tissue stem cells and their immediate descendants can be deleted by apoptosis in response to damage stimuli much less severe than those required to kill other members of the same lineage, if indeed the damage is intrinsically lethal to such cells at all⁶. The *Drosophila* gene *reaper* is a good example of threshold setting: in its absence, the resistance of *Drosophila* embryos to cell death after ionizing radiation is enhanced about 1,000-fold¹³. Indeed the general suicidal tendency of injured stem cells is a testament to the extreme measures adopted to counter the threat posed by progenitors that might have acquired a flawed genome. Failure to initiate apoptosis in response to DNA injury of various types is associated with the appearance of cells with a mutation prevalence one or two orders of magnitude above background^{14,15}. How, then, is DNA damage identified and linked to the apoptosis programme?

Molecular anatomy of a DNA injury response

The eukaryotic strategy to deal with damaged DNA can be split into three components: the recognition of injured DNA, a period of damage assessment (enforced by checkpoints), and the implementation of the appropriate response (DNA repair or cell death). These procedures are not activated in a simple linear fashion, because damage recognition elicits multiple synchronous signals that can trigger both repair and apoptotic processes. Checkpoints have a critical role in the damage response system as they provide an opportunity to monitor the appropriateness of suicide over repair. Checkpoints establish relations between cellular processes so that the execution of one process is contingent on the successful completion of an earlier unrelated activity¹⁶. The checkpoint to oversee the accurate replication of the genome before allowing cell division is an example. In the

context of DNA damage, checkpoints erect barriers to prevent the perpetuation of injured genomes. These can be lifted once the cell has recovered. Occasionally, mutations affect the checkpoint genes themselves. The consequent loss of synchronous quality control can have disastrous results, as seen in the destabilized genomes that are characteristic of cancer¹⁷. This was the threat that early metazoans countered by weaving the apoptosis programme into the web of their checkpoint controls. The existence of multiple points of contact between the apoptotic and checkpoint programmes might explain the heterogeneity of downstream events in the DNA damage response. These mixed signals might compel a cell to die even though DNA repair machines have been successfully engaged¹⁸.

The figure in Box 1 illustrates the major DNA repair options for a mammalian cell. In some cases large complexes of proteins must sequentially assemble over the lesion. This raises the critical question of how DNA damage detectors should be distributed in a manner that allows them to survey the entire genome. Although the ‘active’ nucleotide excision repair (NER) repairosome can tether itself to complexes that naturally navigate the DNA thread, not all repair processes are tied to transcription or replication. An attractive solution would be to corral repair proteins at various nuclear foci for release under conditions of genotoxic stress. One example of this in simple eukaryotes is the discharge of a damage repair protein and chromatin modifiers from yeast telomeres after genotoxic treatment¹⁹. Telomeres are repetitive DNA sequences protected by densely compact chromatin and are particularly appropriate sites in which to sequester detection and repair proteins. Tethered to nuclear pore complexes, yeast telomeres maintain a pool of repair proteins just beneath the nuclear envelope²⁰. A damage-induced flux of repair proteins from them might even provide a useful gauge for the severity of a particular DNA injury.

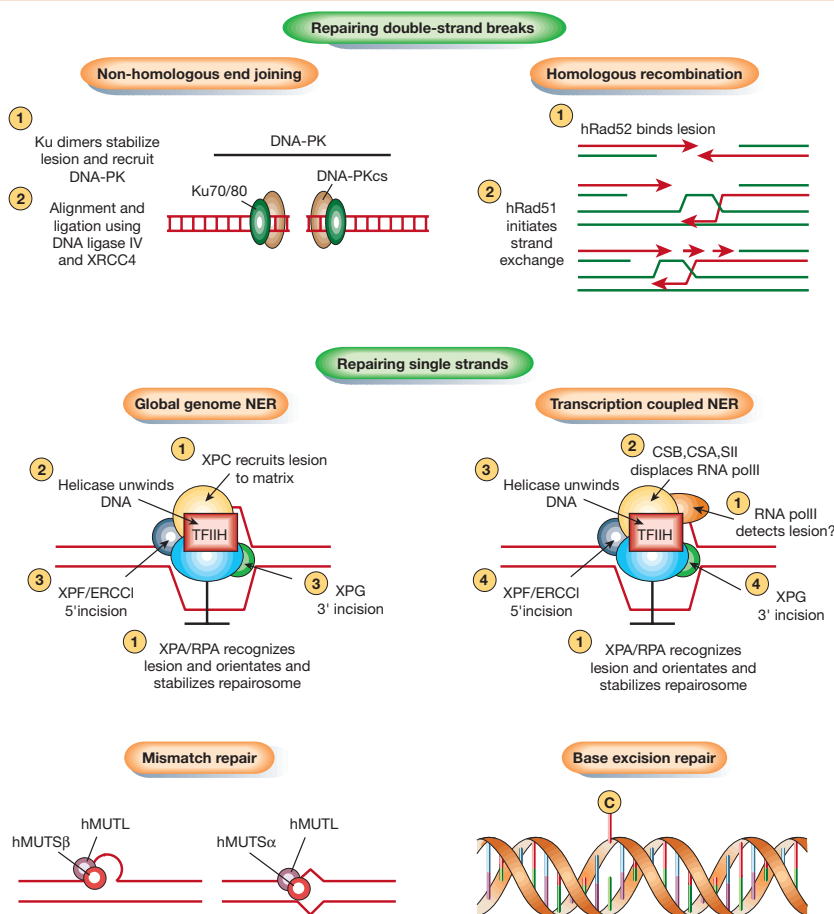
In a striking correlation, the protein components of mammalian telomeres also include DNA repair proteins²¹. A unifying explanation for the tendency of repair proteins to dock at telomeres could be that they regard the ends of the chromosome as a double-strand break (DSB), albeit a naturally occurring one²². Other, naturally occurring, ‘benign’ DSBs use DNA repair proteins for processes such as immune gene *V(D)J* recombination²³. Similarly, hoarding of repair proteins at telomeres might represent a shrewd mechanism for optimizing

Box 1

DNA repair mechanisms

Double-strand break (DSB) repair. A DSB is potentially lethal. Two competing repair processes called homologous recombination and non-homologous end-joining (NHEJ) target DSBs⁷⁵. Homologous recombination uses a sister chromatid or homologue to patch up the damage, whereas NHEJ is less accurate and simply joins DNA ends together. Variations of each process exist, most importantly in the use of conservative or non-conservative homologous recombination, which, as the name suggests, have different mutagenic potentials. NHEJ and homologous recombination are often described as the dominant repair pathways for mammals and yeast respectively. Despite its inaccuracy, mammals seem to favour NHEJ as their repeat-ridden genomes make sequence alignment tricky. But it is now known that vertebrates are also proficient at homologous recombination⁷⁶, prompting a major reassessment of the value of this process to mammalian repair. The mechanics of NHEJ entails the binding of Ku heterodimers to DNA breaks, protecting them from degradation and stabilizing the lesion. Ku then recruits the catalytic subunit of DNA-PK (DNA-PKcs) to activate the DNA-PK holoenzyme. The formation of this activated nucleoprotein complex promotes rejoining by a DNA ligase IV–XRCC4 heterodimer (XRCC4 denotes the X-ray cross-complementation group containing a deletion of the XRCC4 gene product). For the particular case of homologous recombination shown, hRAD51 is recruited to the DNA break, followed by invasion of the intact sister chromatid by hRAD51 to generate a recombination intermediate. As the sister chromatid acts as a template, repair must take place in late S or the G2 phase of the cell cycle. The breast-cancer susceptibility gene *BRCA1* product co-localizes with hRAD51 and promotes homologous recombination, perhaps to discourage the less accurate NHEJ process. There is also considerable interest in the Mre11–Rad50–NBS1 complex, which performs the nucleolytic processing of DSBs and is also implicated in cell-cycle checkpoints through ATM.

Single-strand repair. Single-strand repair is determined by the site and nature of the break. Nucleotide excision repair (NER) is used to excise bulky lesions, such as pyrimidine dimers, that distort the DNA helix. Two NER machines repair the inactive (the global mechanism) and active portion of the genome; RNA polymerases have a major role in the latter. For either complex, lesion recognition is followed by excision of the damaged DNA (steps 1–3 and 1–4 in the respective panels) so that re-replication can occur. Mismatch repair (MMR) detects several types of single-base mismatches in addition to more complicated loops or deletions⁷⁷. Current interest in this process derives from the identification of defective MMR genes as the causative agents of hereditary non-polyposis cancer. Various combinations of hMUTS and hMUTL heterodimers recognize each class of lesion to recruit repairosomes. ATP hydrolysis facilitates either translocation/looping of the DNA or the conversion of hMUTS to a sliding clamp that activates and recruits repair proteins including hMUTL complexes, polymerases- δ/ϵ , exonucleases and replication factors. Finally, base excision repair removes small lesions such as alkylated and methylated bases. This is an ancient repair process that counteracts the natural instabilities of DNA as well as those posed by environmental genotoxins⁷⁸. In the example shown, the damaged base is literally swung out of the helix and into the 'pocket' of a correcting enzyme (yellow ball), which snips it from the helix. The abasic site can be processed by APE1 endonuclease before DNA polymerase- β inserts the correct nucleotide and XCC1/ligase III seals the nick. These proteins may be orientated on PARP. 'Short patch repair' is used in this instance, although 'long patch repair' is available for gaps of two to eight nucleotides.



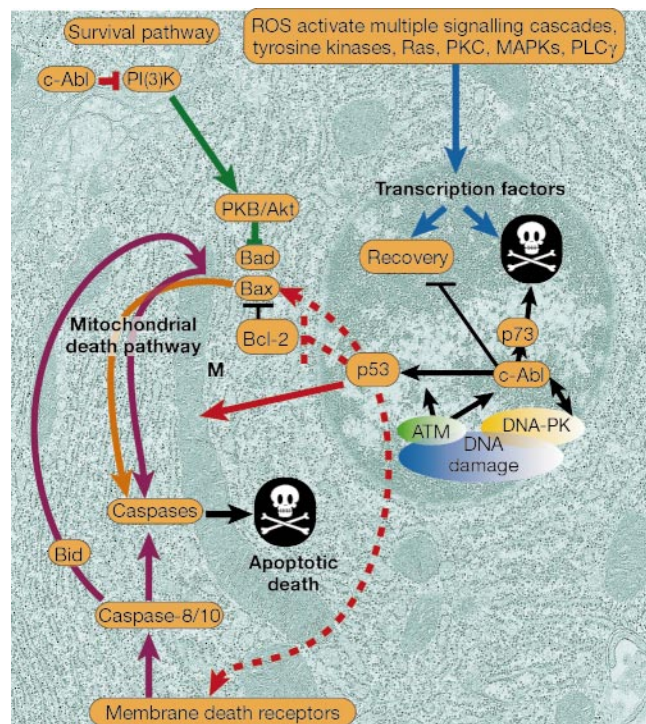
telomeric maintenance. As telomeres shorten with age, the subsequent exposure of chromosome ends can trigger their end-to-end ligation, which is a catastrophic outcome for the cell and its progeny. A checkpoint that forces cells to senesce or undergo apoptosis when telomeres become critically short is required to prevent this occurrence²⁴. One such checkpoint activator, sensitive to the presence of free double-stranded DNA ends, is ATM (for ataxia telangiectasia mutated)²⁵.

The ATM family of DNA damage sensors

ATM is one of a remarkable group of PI(3)K-related kinases that also includes DNA-PKcs (the catalytic subunit of DNA-dependent protein kinase)²⁶ and ATR (ataxia telangiectasia Rad3 related)²⁷. These proteins

are all crucial in detecting the most lethal type of DNA damage, the DSB. ATM encodes a protein with a relative molecular mass of ~350,000 (M_r 350K) containing a DNA-binding domain and a PI(3)K catalytic domain. Atomic force micrographs provide compelling evidence that ATM and DNA-PK bind directly to free DNA ends²⁸. Having done so, these kinases catalyse phosphorylation cascades to transmit damage signals to checkpoints and repair proteins. With the appropriate kinetics, such cascades can operate as sensitive molecular switches²⁹. Exploring this issue further, it is predicted that instabilities in phosphorylation–dephosphorylation cycles could provide the core mechanism of a G2/M checkpoint³⁰. The attractiveness of this model is its inherent ability to ratify each component of the system before proceeding, one of

Figure 2 Apoptotic and survival pathways. Radiation damage triggers multiple stress and apoptotic pathways dependent on the cell type involved. Stress signals generated outside the nucleus include activated mitogen-activated protein kinase (MAPK) cascades (extracellular signal-regulated protein kinase (ERK), JNK and p38) and protein kinase C (PKC) (blue arrow). Transcription factors are an important target of MAPK cascades: ERK activation tends to favour survival, whereas JNK activation assists cell death. The redox-sensitive transcription factor NF- κ B also translocates to the nucleus after its activation by reactive oxygen species (ROS). The activity of this transcription factor is generally associated with protection from apoptosis. Expression levels of several membrane death receptors might be augmented by stabilized p53 (red dashes). These outcompete decoy receptors, leading to the activation of caspases and an additional caspase-dependent pathway that loops through the mitochondrion (M) via Bid. Stabilized p53 also increases the concentration of Bax while diminishing the level of Bcl-2, thus favouring the disruption of mitochondrial membranes and, ultimately, the activation of caspases. The non-receptor tyrosine kinase, c-Abl, has dual roles in the cytoplasm and in the nucleus. The nuclear version is activated by ATM and can stabilize p53. Several pro-apoptotic activities have been suggested for c-Abl, although it is probably fair to say that many of these are still speculative. An important survival pathway (green arrow) is the protein-kinase-B-mediated inactivation of Bad, which is inhibited by cytoplasmic c-Abl.



the central tenets of checkpoint control. In a startling estimate for the sensitivity of these damage detection systems it has been calculated that a single DSB can trigger the arrest of the cell cycle³¹. But why are such large kinases, each with a M_r >250K, required to detect DNA damage? One possibility is that these proteins could provide a platform over which other detectors and repair proteins can assemble.

ATM, ATR and DNA-PK act as checkpoint sensors that signal to both cell-cycle and apoptosis machines. Figure 1 outlines the various cell-cycle arrests that can be instigated after the activation of ATM. Given that DNA can be damaged at any point of the cell cycle, multiple checkpoints are required to ensure a comprehensive arrest strategy for each phase³². However, the flaw in this system is that the overall control of these pathways rests in the hands of relatively few sensor molecules so that a single protein often polices multiple phase transitions. For example, p21 can arrest the cell at G1/S or in situations of abortive mitosis³³. Similarly, ATM can signal to checkpoint arrests throughout the cell cycle³⁴. Tracking from G1 through the cell cycle (Fig. 1) it can be seen that p21, a potent inhibitor of cyclin-dependent kinases, is transactivated by p53 and p73, although p73 has not yet been shown to be a genuine tumour suppressor³⁵. Damage incurred while DNA is replicating during S phase prevents fresh origins of replication from being fired. At this stage the crucial components are the mammalian CHK1 and CHK2 proteins, which, after phosphorylation by ATM, can inhibit the phosphatases required for G1/S and G2/M progression³². The checkpoint strategy used at the G2/M phase of the cycle provides a textbook example of how the separation of an enzyme from its substrate can block unwanted activity. In this instance the CDC25C phosphatase is prevented from activating the cdc2–cyclin B complex that is required for entry to mitosis³⁶. The recent demonstration that ATM links p95/Nbs1 to an S-phase checkpoint³⁷ indicates that the checkpoints described here might represent only a subset of those that operate through ATM. They do, however, illustrate how checkpoint molecules that detect DNA damage can force injured cells to engage in cycle arrest and repair their DNA. There are already examples of other, equally specific molecules (for example, the mismatch repair protein MSH-2) that can detect very different types of DNA damage (nucleotide mismatch or inappropriate methylation) to force similar outcomes¹⁴. However, although the proteins involved in the initial recognition and repair of DNA damage have

been known for some time, the means by which they induce the terminal events of apoptosis are not yet clear.

p53 signals to apoptosis effector pathways

p53 provides one well-worked example of how the decision between apoptosis and other fates can be made at checkpoints activated by DNA damage³⁸. Checkpoint activation, involving ATM and other recognition molecules, leads to p53 phosphorylation, which alters its conformation and greatly increases its stability. Several amino-terminal serines are consistently phosphorylated after radiation-induced DNA damage, and there is some specificity of mechanism. For example, phosphorylation by ATM preferentially occurs at Ser 15, whereas DNA-PK modifies Ser 15 and Ser 37 (details of p53 modifications are reviewed in depth elsewhere³⁹). For most replicative cell populations, p53 levels increase within minutes of DNA damage and the first apoptotic events occur within a few hours. No early death is seen within tissues engineered to have no p53 (refs 40, 41). How, then, does the activation of p53 by DNA damage lead to the initiation of apoptosis? Several cell-cycle regulators are induced by p53, for example p21, GADD45 and members of the 14-3-3 family. Other induced proteins include Bax, CD95, DR5 (a receptor for the death ligand TRAIL)⁴² and (in *Drosophila*) Rpr (ref. 43), which are all classical members of the core apoptosis pathways (red dashes in Fig. 2). However, the significance of these inductions remains somewhat obscure, as some cells from *bax*^{-/-} and *gld* (CD95-inactive) mice show normal radiation sensitivity⁴⁴. Moreover, CD95 induction is dependent on a p53-response element in the first intron (reassuringly conserved between mammalian species) that is activated equally by wild-type p53 and point mutants that are inactive in initiating apoptosis⁴⁵. A further important p53-induced protein is MDM2. This escorts p53 from the nucleus and targets it for proteasomal degradation, thus ensuring that the p53 signal is transient and carefully controlled.

The advent of microarray and other genome-wide technologies has drawn attention to scores of newly transcribed molecules that might transmit the p53 signal to the apoptotic machinery⁴⁶. One convincing newcomer is PERP, a four-span plasma membrane protein with similarity to the PMP-22/Gas3 family⁴⁷. This transcript is associated exclusively with the apoptotic rather than the cycle-arrest

functions of p53. A separate mechanism is suggested by the induction of MIC-1 (a secreted transforming growth factor- β -like cytokine)⁴⁸ and IGF-BP3 (a secreted binding protein for the survival factor IGF-1). These proteins could conceivably promote apoptosis through alteration of the cellular microenvironment. Perhaps the most remarkable challenge to conventional paradigms is the observation that a proportion of stabilized p53 finds its way on to mitochondrial membranes (solid red line in Fig. 2)⁴⁹. Mitochondrial targeting seems to occur only in the context of cells within which the induction of p53 promotes death rather than cell-cycle arrest. Moreover, variants and wild-type p53 engineered to target mitochondria in the absence of any nuclear signal can induce apoptosis. p53 also binds to centrosomes and other components of the mitotic spindle⁵⁰. This invites the question of whether these sites also nucleate primed apoptosomes and, if so, whether they are activated by binding p53.

E2F-1 activity and apoptosis

A second candidate linking DNA damage to apoptosis is the transcription factor E2F-1. This protein is released from the pocket of Rb as it becomes phosphorylated during cell-cycle progression through G1. Concomitant with the induction of the immediate early genes of DNA replication (including, among many others, the proto-oncogene *c-myc*), E2F-1 heterodimerizes with DP-1 (ref. 51). It is now known that both E2F-1 and p53 lie within a DNA damage pathway⁵² and become stabilized after exposure to ionizing radiation or ultraviolet C radiation. Like p53, E2F-1 is bound and inactivated by hDM2 (the human version of MDM2), at the same time releasing DP-1 to the nucleus. Moreover, the expression of E2F-1 can initiate apoptosis, even in a p53-null background. Thus, hDM2 can act as a survival factor, independently of its interaction with p53, through its ability to bind and destabilize E2F-1. In a new development, two groups now place E2F-1 and p73 in an apoptosis pathway, providing a mechanism for the E2F-1-mediated killing that can occur in the absence of p53 (refs 53, 54). One protocol exploited receptor-mediated 'hyper' stimulation to kill T cells. Death coincided with the induction of p73 and the peak of E2F-1's transcriptional activity — which is at the late G1 or S phase. Given appropriate circumstances, the association of both E2F-1 (ref. 55) and c-Myc (ref. 56) with apoptosis rather than cell proliferation suggests that entry to a replicative (or pre-replicative) state is somehow necessary to initiate apoptosis in cells bearing damage to DNA. This concept would readily fit with long-established observations on the role of the Rb-binding adenoviral protein E1A in the response of fibroblasts to ionizing

radiation. The irradiation of primary fibroblasts leads to pre-replicative cell-cycle arrest, but in fibroblasts transfected with E1A, the Rb protein is silenced and E2F is released from its pocket so that the cells respond to an identical injury by entering apoptosis⁵⁷.

c-Abl activity and apoptosis

A third substrate of ATM phosphorylation after DNA injury is the proto-oncoprotein c-Abl. c-Abl is a Src-like tyrosine kinase with an unusual carboxy-terminal domain that contains nuclear localization signals and DNA-binding sites^{58,59}. In accordance with its distribution to both the nucleus and the cytoplasm, immunoprecipitation data suggest that it binds DNA-PK, ATM, Rad51, Rb, p53, p73 and perhaps other proteins⁶⁰. After damage to DNA by ionizing radiation, c-Abl might be activated by phosphorylation through an ATM-dependent mechanism to enhance its kinase activity. DNA-PK also phosphorylates c-Abl, which in turn phosphorylates DNA-PKcs in a feedback mechanism that causes it to dissociate from Ku (ref. 59). Theoretically, therefore, c-Abl activation also contributes to turning off a signal at the heart of damage detection. The cycle arrest and apoptosis that are normally induced by ionizing radiation are prevented in cells that lack c-Abl or possess only a kinase-dead mutant. Although c-Abl is known to be an ATM substrate and can interact with many of the nucleoproteins concerned with the cellular response to DNA injury, the significance of most of its reactions is not yet clear⁵⁹.

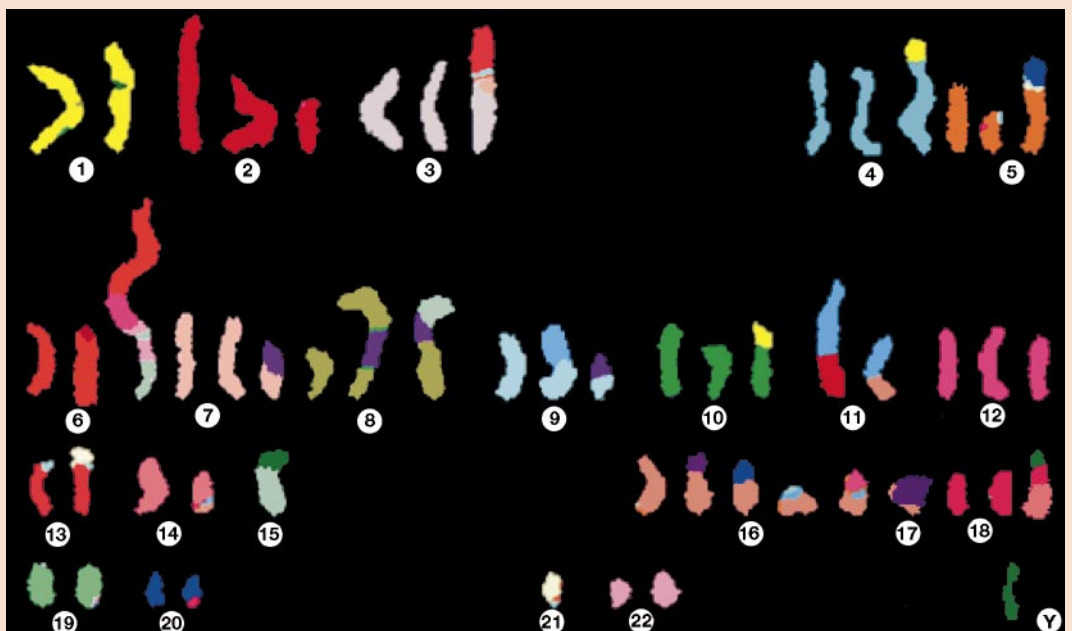
The question arises of why the signals that couple DNA damage to the apoptosis machinery need to be so redundant and complex. One possible answer derives from the observation that many of the signals favouring death can be overridden. Presumably the many stimuli arriving in the injured cell define a threshold for apoptosis that can vary with time. The final decision to initiate apoptosis rather than cell-cycle arrest or a failure to respond by either route is likely to be conditioned by the magnitude and duration of the damage stimulus. It will also reflect the damaged cell's replicative status, its recent history as demonstrated by the availability of MDM2 or CD95, and even its position, because the local growth factor environment expresses proximity to neighbouring cells and to basement membrane.

A nuclear apoptosome

Jeffrey Nickerson in 1998 remarked: "There are, however, two properties of tumors that are fundamental and that define some tumors as malignant. These are, first, alterations in the architecture of cells and tissues and, second, genetic instability. Both of these

Figure 3 Spectral karyotyping.

This metaphase image of the breast cancer cell line MDA-MB-361 (courtesy of J. Davidson, Department of Pathology, University of Cambridge) was obtained by 24-colour fluorescence *in situ* hybridization, with spectral imaging as described by Schrock *et al.*⁷². Each chromosome is labelled with a different combination of fluorescent dyes and the final image is interpreted by software that colours each pixel to show which chromosome is most likely to be present at that point.



hallmarks of cancer may be addressed by examination of the nuclear structure.⁶¹ In fact, it seems that both are intimately connected. Repair systems must contend with the complex topology of DNA, probably by anchoring it to the nuclear matrix. In addition, massive nuclear complexes are known to choreograph multiple nuclear functions. Indeed, there is accumulating evidence that the nucleus is a burgeoning mass of these supercomplexes, several of which are heavily implicated in apoptosis and DNA repair.^{62–64} One is the PML body, which takes its name from the cancer (promyelocytic leukaemia) that disrupts its structure^{65,66}. PML bodies (also called promyelocytic oncogenic domains) are nucleated by multimers of the PML protein. PML procures a large number of nucleoproteins, crucial to almost the entire range of nuclear functions, and stores them in PML bodies. The mode of this recruitment is largely unknown, although modification by the ubiquitin-related modifier (SUMO-1) seems to be one mechanism⁶⁶. PML bodies contain the Nijmegen breakage syndrome disease protein (p95/Nbs1), which assists in the repair of DSBs⁶⁷. PML might also act in concert with DAXX (a transcriptional repressor) to potentiate apoptosis⁶⁸, a theory supported by the resistance observed in PML-deficient systems from multiple apoptotic stimuli⁶³. Sorties of damage and repair proteins from PML bodies are likely to occur when these proteins are in demand, and the natural fluctuations in the size and number of PML bodies could reflect their servicing of various nuclear machines.

The sequestration of molecules by scaffold proteins is a familiar concept in the field of protein signalling and imposes order and substrate specificity on proteins that are common to several pathways. The existence of such regulatory supercomplexes within the nucleus would be especially prudent given that crucial repair nucleases cannot be allowed to diffuse freely. A supercomplex of tumour suppressors and DNA damage and repair proteins called BASC (for 'BRCA1-associated genome surveillance complex') has recently been described⁶². BRCA1 is another enormous protein (1,863 residues) that, by virtue of expressing a BRCT (BRCA1 C-terminal) domain, is part of a superfamily of DNA damage and cell-cycle checkpoint proteins. BASC bodies might be multiple aggregates of repairosomes that can be remodelled (perhaps by ubiquitinating BRCA1's RING finger domains) to suit each type of damage detected. It will be interesting to determine whether BASC complexes act as distributors or as platforms for DNA repair engines and to examine the importance of BRCA1 as a scaffold for these complexes. BRCA1's latest incarnation as part of the human version of the SWI/SNF complex (which remodels nucleosomes) is equally provocative and might explain how BRCA1 regulates transcription⁶⁹. Significantly, many of the cancer-associated exon 11 deletions of BRCA1 also negate its ability to associate with the SWI/SNF complex.

Cancer is associated with gross alterations to the organization of the nuclear matrix and is therefore likely to affect both DNA metabolism and subnuclear organization. In fact, many cancers are typified by complex recombination, often involving three or more chromosomes. Chromosomes reside in distinct territories of the interphase nucleus and we might expect recombination events to be more likely between adjacent chromosomes. The loss of p21^{WAF1/CIP} alone is sufficient to reorganize these domains⁷⁰, so it is possible that a cell lacking a critical checkpoint gene might have shuffled the position of its chromosome territories to favour such events. As the signature lesions of cancer are often rapidly immersed in successive waves of mutations, sometimes as many as 10³–10⁵ per tumour cell, it can be extremely difficult to identify the founder lesions⁷¹. The spectral karyotyping technique (Fig. 3) might help us to uncover the recurrent patterns of chromosomal aberrations that characterize these lesions⁷² and the 'master genes' that control them.

Future directions

DNA damage and apoptosis are both fast-growing fields and our current knowledge might well be only scratching the surface of what awaits. However, a growing understanding of these processes is already

paying dividends. During preparation of this manuscript, the success of the first phase II clinical trial to combat cancer by using a modified virus (Onyx015) was announced. Onyx015 can only replicate in, and ultimately kill, p53-null cells⁷³. Early concerns about replication in p53-positive tumours were allayed when it emerged that such tumours often lack other elements in the p53 response pathway. Away from the clinic, there have been considerable successes in our understanding of the cell biology of DNA damage responses. Many of these have arisen from fresh initiatives to study nuclear organization. Confocal and other imaging techniques have been trained on this most enigmatic organelle in a renewed attempt to understand its organization. The reward is that we can now appreciate some of the unique stratagems devised by the nucleus to coordinate its work. Given that compartmentation and supercomplexes provide some of the answer, it seems only a matter of time before a nuclear apoptosome is described. A fresh impetus to these studies comes from the revelation that PML bodies might also control the entry to cellular senescence by regulating p53 acetylation⁷⁴. We should also ask about the repair lesion itself: How is chromatin remodelled during repair? Does it re-organize to stabilize the lesion or to block checkpoint signals²²? Does chromatin exclude some proteins and attract others to repair sites? These are all crucial questions because close cooperation must exist between nucleosome remodelling and repair proteins to ensure access of the latter and facilitate repair. Delays might allow the levels of stabilized p53 to creep up and increase the chance of activating apoptosis. This leads us to the central problem of how this cell fate is sometimes sanctioned over any other. Identifying new damage-induced transcripts in cultured cells will go some way to answering this, but at some point we shall have to resolve the responses of cells in their authentic tissue microenvironment. □

- Nagata, S. & Golstein, P. The Fas death factor. *Science* **267**, 1449–1456 (1995).
- Kroemer, G. & Reed, J. C. Mitochondrial control of cell death. *Nature Med.* **6**, 513–519 (2000).
- Korsmeyer, S. J. BCL-2 gene family and the regulation of programmed cell death. *Cancer Res.* **59**, 1693–1700 (1999).
- Friedberg, E., Walker, G. & Siede, W. *DNA Repair and Mutagenesis* 108–133 (ASM Press, Washington DC, 1995).
- MacCallum, D. E. *et al.* The p53 response to ionising radiation in adult and developing murine tissues. *Oncogene* **13**, 2575–2587 (1996).
- Bach, S. P., Renahan, A. G. & Potten, C. S. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* **21**, 469–476 (2000).
- Lowe, S. W., Schmitt, E. M., Smith, S. W., Osborne, B. A. & Jacks, T. p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature* **362**, 847–849 (1993).
- Clarke, A. R. *et al.* Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature* **362**, 849–852 (1993).
- Matsuyama, S., Nouraini, S. & Reed, J. C. Yeast as a tool for apoptosis research. *Curr. Opin. Microbiol.* **2**, 618–623 (1999).
- Aravind, L., Dixit, V. M. & Koonin, E. V. The domains of death: evolution of the apoptosis machinery. *Trends Biochem. Sci.* **24**, 47–53 (1999).
- Oshima, H., Oshima, M., Kobayashi, M., Tsutsumi, M. & Taketo, M. M. Morphological and molecular processes of polyp formation in Apc(delta716) knockout mice. *Cancer Res.* **57**, 1644–1649 (1997).
- Pieper, A. A., Verma, A., Zhang, J. & Snyder, S. H. Poly (ADP-ribose) polymerase, nitric oxide and cell death. *Trends Pharmacol. Sci.* **20**, 171–181 (1999).
- White, K. *et al.* Genetic control of programmed cell death in *Drosophila*. *Science* **264**, 677–683 (1994).
- Toft, N. J. *et al.* Msh2 status modulates both apoptosis and mutation frequency in the murine small intestine. *Proc. Natl Acad. Sci. USA* **96**, 3911–3915 (1999).
- Griffiths, S. D. *et al.* Absence of p53 permits propagation of mutant cells following genotoxic damage. *Oncogene* **14**, 523–531 (1997).
- Hartwell, L. H. & Weinert, T. A. Checkpoints: controls that ensure the order of cell cycle events. *Science* **246**, 629–634 (1989).
- Hanahan, D. & Weinberg, R. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Van Sloun, P. P. *et al.* The role of nucleotide excision repair in protecting embryonic stem cells from genotoxic effects of UV-induced DNA damage. *Nucleic Acids Res.* **27**, 3276–3282 (1999).
- Martin, S., Laroche, T., Suka, N., Grunstein, M. & Gasser, S. Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell* **97**, 621–633 (1999).
- Galy, V. *et al.* Nuclear pore complexes in the organization of silent telomeric chromatin. *Nature* **403**, 108–112 (2000).
- Hsu, H. L., Gilley, D., Blackburn, E. H. & Chen, D. J. Ku is associated with the telomere in mammals. *Proc. Natl Acad. Sci. USA* **96**, 12454–12458 (1999).
- Gasser, S. A sense of the end. *Science* **288**, 1377–1379 (2000).
- Lieber, M. Pathological and physiological double-strand breaks. Roles in cancer, aging and the immune system. *Am. J. Pathol.* **153**, 1323–1332 (1998).
- Artandi, S. E. & DePinho, R. A. A critical role for telomeres in suppressing and facilitating carcinogenesis. *Curr. Opin. Genet. Dev.* **10**, 39–46 (2000).
- Rotman, G. & Shiloh, Y. ATM: from gene to function. *Hum. Mol. Genet.* **7**, 1555–1563 (1998).
- Smith, G. C. M. & Jackson, S. P. The DNA-dependent protein kinase. *Genes Dev.* **13**, 916–934 (1999).
- Hoekstra, M. F. Responses to DNA damage and regulation of cell cycle checkpoints by the ATM protein kinase family. *Curr. Opin. Genet. Dev.* **7**, 170–175 (1997).
- Smith, G. C. *et al.* Purification and DNA binding properties of the ataxia-telangiectasia gene product

- ATM. *Proc. Natl Acad. Sci. USA* **96**, 11134–11139 (1999).
29. Goldbeter, A. & Koshland, D. E. Jr An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78**, 6840–6844 (1981).
30. Aguda, B. Instabilities in phosphorylation–dephosphorylation cascades and cell cycle checkpoints. *Oncogene* **18**, 2846–2851 (1998).
31. Huang, L. C., Clarkin, K. C. & Wahl, G. M. Sensitivity and selectivity of the DNA damage sensor responsible for activating p53-dependent G1 arrest. *Proc. Natl Acad. Sci. USA* **93**, 4827–4832 (1996).
32. Dasika, G. *et al.* DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. *Oncogene* **18**, 7883–7899 (1999).
33. Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497–1501 (1998).
34. Pandita, T. K. *et al.* Ionizing radiation activates the ATM kinase throughout the cell cycle. *Oncogene* **19**, 1386–1391 (2000).
35. Lohrum, M. & Vousden, K. Regulation and function of the p53-related proteins: same family, different rules. *Trends Cell Biol.* **10**, 197–202 (2000).
36. Chan, T., Hermeking, H., Lengauer, C., Kinzler, K. & Vogelstein, B. 14-3-3 σ is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**, 616–620 (1999).
37. Lim, D. S. *et al.* ATM phosphorylates p53/nbs1 in an S-phase checkpoint pathway. *Nature* **404**, 613–617 (2000).
38. Sionov, R. V. & Haupt, Y. The cellular response to p53: the decision between life and death. *Oncogene* **18**, 6145–6157 (1999).
39. Meek, D. Mechanisms of switching on p53: a role for covalent modification? *Oncogene* **18**, 7666–7675 (1999).
40. Clarke, A. R., Gledhill, S., Hooper, M. L., Bird, C. C. & Wyllie, A. H. p53 dependence of early apoptotic and proliferative responses within the mouse intestinal epithelium following gamma-irradiation. *Oncogene* **9**, 1767–1773 (1994).
41. Merritt, A. J. *et al.* The role of p53 in spontaneous and radiation-induced apoptosis in the gastrointestinal tract of normal and p53-deficient mice. *Cancer Res.* **54**, 614–617 (1994).
42. Wu, G. S. *et al.* Induction of the TRAIL receptor KILLER/DR5 in p53-dependent apoptosis but not growth arrest. *Oncogene* **18**, 6411–6418 (1999).
43. Brodsky, M. H. *et al.* *Drosophila* p53 binds a damage response element at the reaper locus. *Cell* **101**, 103–113 (2000).
44. O'Connor, L., Harris, A. W. & Strasser, A. CD95 (Fas/APO-1) and p53 signal apoptosis independently in diverse cell types. *Cancer Res.* **60**, 1217–1220 (2000).
45. Munsch, D. *et al.* Human and mouse Fas (APO-1/CD95) death receptor genes each contain a p53-responsive element that is activated by p53 mutants unable to induce apoptosis. *J. Biol. Chem.* **275**, 3867–3872 (2000).
46. Zhao, R. *et al.* Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev.* **14**, 981–993 (2000).
47. Attardi, L. D. *et al.* PERP, an apoptosis-associated target of p53, is a novel member of the PMP-22/gas3 family. *Genes Dev.* **14**, 704–718 (2000).
48. Kannan, K., Amariglio, N., Rechavi, G. & Givol, D. Profile of gene expression regulated by induced p53: connection to the TGF- β family. *FEBS Lett.* **470**, 77–82 (2000).
49. Marchenko, N. D., Zaika, A. & Moll, U. M. Death signal-induced localization of p53 protein to mitochondria. A potential role in apoptotic signaling. *J. Biol. Chem.* **275**, 16202–16212 (2000).
50. Morris, V. B., Brammall, J., Noble, J. & Reddel, R. p53 localizes to the centrosomes and spindles of mitotic cells in the embryonic chick epiblast, human cell lines, and a human primary culture: an immunofluorescence study. *Exp. Cell Res.* **256**, 122–130 (2000).
51. Loughran, O. & La Thangue, N. B. Apoptotic and growth-promoting activity of E2F modulated by MDM2. *Mol. Cell. Biol.* **20**, 2186–2197 (2000).
52. Blattner, C., Sparks, A. & Lane, D. Transcription factor E2F-1 is upregulated in response to DNA damage in a manner analogous to that of p53. *Mol. Cell. Biol.* **19**, 3704–3713 (1999).
53. Lissy, N. A., Davis, P. K., Irwin, M., Kaelin, W. G. & Dowdy, S. F. A common E2F-1 and p73 pathway mediates cell death induced by TCR activation. *Nature* **407**, 642–645 (2000).
54. Irwin, M. *et al.* Role for the p53 homologue p73 in E2F-1-induced apoptosis. *Nature* **407**, 645–648 (2000).
55. Wu, X. & Levine, A. J. p53 and E2F-1 co-operate to mediate apoptosis. *Proc. Natl Acad. Sci. USA* **91**, 3602–3606 (1994).
56. Evan, G. I. *et al.* Induction of apoptosis in fibroblasts by c-myc protein. *Cell* **69**, 119–128 (1992).
57. Lowe, S. W., Jacks, T., Housman, D. E. & Ruley, H. E. Abrogation of oncogene-associated apoptosis allows transformation of p53-deficient cells. *Proc. Natl Acad. Sci. USA* **91**, 2026–2030 (1994).
58. Van Etten, R. A. Cycling, stressed-out and nervous: cellular functions of c-Abl. *Trends Cell Biol.* **9**, 179–186 (1999).
59. Kharbanda, S., Yuan, Z. M., Weichselbaum, R. & Kufe, D. Determination of cell fate by c-Abl activation in the response to DNA damage. *Oncogene* **17**, 3309–3318 (1998).
60. Shaul, Y. c-Abl: activation and nuclear targets. *Cell Death Differ.* **7**, 10–16 (2000).
61. Nickerson, J. Nuclear dreams: the malignant alteration of nuclear architecture. *J. Cell. Biochem.* **70**, 172–180 (1998).
62. Wang, Y. *et al.* BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927–939 (2000).
63. Wang, Z. G. *et al.* PML is essential for multiple apoptotic pathways. *Nature Genet.* **20**, 266–272 (1998).
64. Zhong, S. *et al.* A role for PML and the nuclear body in genomic stability. *Oncogene* **18**, 7941–7947 (1999).
65. Hodges, M., Tissot, C., Howe, K., Grimwade, D. & Freemont, P. S. Structure, organization, and dynamics of promyelocytic leukemia protein nuclear bodies. *Am. J. Hum. Genet.* **63**, 297–304 (1998).
66. Maul, G. G., Negorev, D., Bell, P. & Ishov, A. M. Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. *J. Struct. Biol.* **129**, 278–287 (2000).
67. Lombard, D. B. & Guarente, L. Nijmegen breakage syndrome disease protein and MRE11 at PML nuclear bodies and meiotic telomeres. *Cancer Res.* **60**, 2331–2334 (2000).
68. Zhong, S. *et al.* Promyelocytic leukemia protein (PML) and Daxx participate in a novel nuclear pathway for apoptosis. *J. Exp. Med.* **191**, 631–640 (2000).
69. Bochar, D. *et al.* BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
70. Linares-Cruz, G. *et al.* p21WAF-1 reorganizes the nucleus in tumor suppression. *Proc. Natl Acad. Sci. USA* **95**, 1131–1135 (1998).
71. Gray, J. & Collins, C. Genome changes and gene expression in human solid tumors. *Carcinogenesis* **21**, 443–452 (2000).
72. Schrock, E. *et al.* Multicolor spectral karyotyping of human chromosomes. *Science* **273**, 494–497 (1996).
73. Anderson, W. Gene therapy scores against cancer. *Nature Med.* **6**, 862–863 (2000).
74. Pearson, M. *et al.* PML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature* **406**, 207–210 (2000).
75. Karran, P. DNA double strand breaks in mammalian cells. *Curr. Opin. Genet. Dev.* **10**, 144–150 (2000).
76. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 5172–5177 (1998).
77. Kolodner, R. & Marsischky, G. Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* **9**, 89–96 (1999).
78. Krokan, H. E., Nilsen, H., Skorpén, F., Otterlei, M. & Slupphaug, G. Base excision repair of DNA in mammalian cells. *FEBS Lett.* **476**, 73–77 (2000).

Corpse clearance defines the meaning of cell death

John Savill* & Valerie Fadok†

* University of Edinburgh/Medical Research Council Centre for Inflammation Research, Department of Clinical and Surgical Sciences (Internal Medicine), Royal Infirmary, Edinburgh EH3 9YW, UK (e-mail: j.savill@ed.ac.uk)

† Program in Cell Biology, Department of Pediatrics, National Jewish Medical and Research Center, D509, 1400 Jackson Street, Denver Colorado 80206, USA

While philosophers seek the meaning of life, cell biologists are becoming ever more interested in the meaning of death. Apoptosis marks unwanted cells with 'eat me' signals that direct recognition, engulfment and degradation by phagocytes. Far from being the end of the story, these clearance events allow scavenger cells to confer meaning upon cell death. But if the phagocytic 'spin doctors' receive or transmit the wrong messages, trouble ensues.

Since the seminal description of cell death by apoptosis¹ it has been clear that the final phase of this programme of cell deletion *in vivo* is swift and safe phagocytosis of intact 'unwanted' cells^{2,3}. For many years the significance of this event has been underestimated. New data indicate that phagocyte clearance of cells dying by apoptosis is much more than mere waste disposal. Instead, the engulfment of dying cells by phagocytes may define the meaning of cell death in higher organisms, particularly when clearance is achieved by 'professional' scavengers of the macrophage line rather than by neighbouring cells acting as 'semi-professional' phagocytes. Depending on the context, the removal of apoptotic cells by phagocytes might suppress inflammation, modulate the macrophage-directed deletion of host cells or invading parasites and critically regulate immune responses (Fig. 1). How can such meaning be conferred on the clearance of dying cells by phagocytes?

Clearance of unwanted self

'Eat me' signals

Two approaches have helped to define 'eat me' signals displayed by apoptotic cells and the phagocyte receptors that recognize such surface changes and mediate the engulfment of dying cells (Fig. 2). First, a basic system for signalling engulfment has been defined by the study of genetic defects in the clearance of cell corpses arising during development of the nematode *Caenorhabditis elegans*⁴⁻⁷ (Fig. 3). Second, the characterization of inhibitors that specifically block the phagocyte recognition of apoptotic cells *in vitro* has implicated a repertoire of mammalian cell-surface and extracellular 'bridging' molecules (Fig. 4) for which roles *in vivo* are now being confirmed in deletion mutants.

Cells undergoing apoptosis can display a number of 'eat-me' flags. Some are relatively well characterized, such as the exposure of phosphatidylserine (PtdSer) normally restricted to the inner-membrane leaflet of the dying cell⁸ (as discussed further below), or changes in surface sugars detected by phagocyte lectins⁹. Other 'eat-me' markers are more poorly defined, such as sites that bind adhesive 'bridging' molecules present in extracellular fluid. These include C1q, the first component of complement, for which a bridging role *in vivo* has been confirmed by elegant studies

demonstrating the defective clearance of apoptotic cells in deletion-mutant mice^{10,11}. Potential bridging roles for other components of the complement cascade such as iC3b (ref. 12), the abundant serum protein β_2 glycoprotein I (β_2 GPI)¹³ and thrombospondin², also need clarification. Moreover, we must characterize how phagocyte scavenger receptors such as CD68 (refs 14,15) recognize apoptotic cell-surface sites that can be masked by antibodies against oxidized low-density lipoproteins¹⁶ and how the immunoglobulin superfamily member ICAM-3 (intercellular adhesion molecule-3) can function as an 'eat-me' flag¹⁷.

The phagocyte's reaction

Although we have a much clearer picture of the mammalian phagocyte receptors capable of mediating the engulfment of dying cells (Fig. 4), we do not yet know what contribution is made to the clearance task by each molecule. Existing data suggest that macrophages might 'tether' dying cells by using phagocyte surface CD14 (ref. 18) or β_2 integrins¹² (which bind the potential bridging complement fragment iC3b), before engaging receptors that drive phagocytosis. Indeed, new data¹⁹ demonstrate that the phagocyte integrin $\alpha_v\beta_5$ may direct the cytoskeletal changes necessary for the phagocyte surface to envelop apoptotic cells by recruiting the CrkII-DOCK 180-Rac1 signalling complex that was first implicated by mutations in the respective *C. elegans* homologues CED-2, CED-5 and CED-10 (Fig. 3), which not only inhibit the clearance of cell corpses but also affect the migration of gonadal distal tip cells in the nematode^{4,7}. Similarly, the mammalian class B scavenger CD36 receptor might interact with the homologue of CED-6 (ref. 20), a signalling adaptor protein⁵. Although the identity of CED-1 is not yet clear, we speculate that this *C. elegans* molecule might be some form of scavenger receptor (Fig. 3).

We also need to clarify the contribution made by subtle and dynamic reorganization of the phagocyte membrane, which is dependent on lipid fluxes similar to those occurring in the dying 'prey'. In mammalian cells these lipid transport events, which include the exposure of PtdSer, are directed by the ABC1 transporter^{21,22}, a homologue of the *C. elegans* protein CED-7, which must function in both scavenger cell and target for efficient engulfment⁶ (Fig. 3). Perhaps these phagocyte membrane changes, potentially driven either by apoptotic cell ligation of phagocyte receptors or by the release of soluble signals from dying cells, explain why

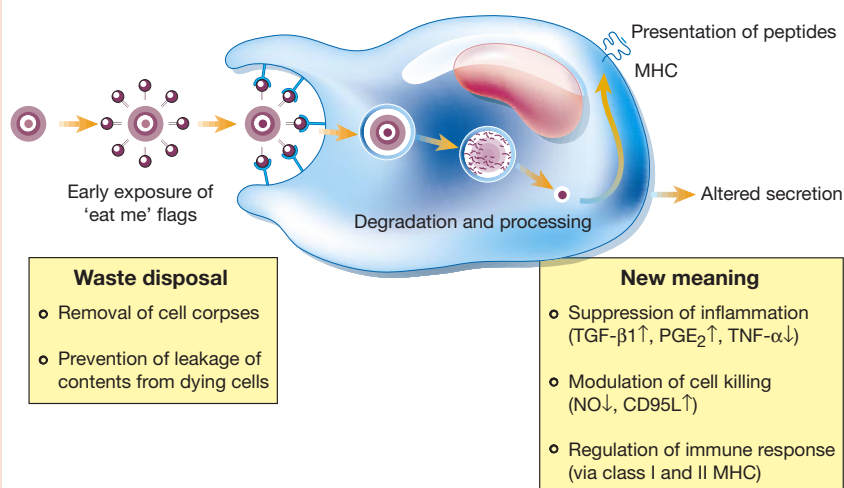


Figure 1 The meaning coded into the phagocytic clearance of cells dying by apoptosis. Until recently, engulfment of apoptotic cells was viewed as protective waste disposal (left). However, recent data reviewed in this article indicate a new meaning (right) determined by phagocyte responses that modulate inflammation, control tissue remodelling by phagocyte-directed cell killing and regulate immune responses. CD95L, CD95 ligand; MHC, major histocompatibility complex; NO, nitric oxide.

ubiquitously expressed cell-surface molecules such as lectins or integrins (like the $\alpha_v\beta_3$ vitronectin receptor) can acquire a specific role in the uptake of apoptotic cells by a wide range of phagocytes, both professional and semi-professional^{2,3,9}.

Nevertheless, other phagocytic receptor recognition mechanisms might have a prominent role only in particular tissues (for example, defective apoptotic cell clearance in C1q-null mice is particularly evident in the kidney¹⁰) or in particular cells. Thus, the expression of CD36 is restricted to macrophages and a few other cell types and yet an important role in the clearance of dying cells during development is demonstrated by defects in a *Drosophila melanogaster* mutant deficient in the CD36 homologue Croquemort²³. Similarly, although healthy mice deficient in the class A scavenger receptor do not demonstrate an obvious defect in the clearance of apoptotic cells from the thymus and other organs²⁴, unpublished work (G. Thomas, N. Platt, S. Gordon and J.S.) points to a defect *in vivo* in the ingestion of apoptotic leukocytes by macrophages during peritoneal inflammation.

Our current understanding of the molecular mechanisms mediating the clearance of apoptotic cells therefore remains poor. The apparent redundancy of these mechanisms could be consistent with the probable importance for health of safely clearing apoptotic cells, but might also reflect how little we know about the circumstances *in vivo* in which various mechanisms are engaged. Indeed, the plethora of 'eat-me' signals, bridging molecules and phagocyte receptors provides a rich substrate from which the clearance event could acquire additional meaning. The potential significance programmed into these molecular mechanisms was suggested by a simple experiment *in vitro*²⁵. Macrophages are crucial both for the clearance of apoptotic cells generated in injured tissue and for host defence against infection by bacteria or protozoa. Normally the ingestion of particles of similar size to these invaders triggers macrophages to secrete molecular mediators capable of initiating protective but potentially injurious inflammatory responses. But the ingestion of large numbers of apoptotic cells failed to elicit this macrophage release of pro-inflammatory mediators²⁵. However, such a release was observed if experimental conditions were deliberately changed so that the recognition mechanisms for 'quiet clearance' were replaced by bridging immunoglobulin and macrophage Fc receptors²⁵. Thus, mechanisms allowing phagocytes to recognize apoptotic cells as 'unwanted self' are special in that they are uncoupled from inflammatory responses. As we discuss in the next section, these molecular mechanisms might have a much wider significance in health and disease.

Suppression of inflammation

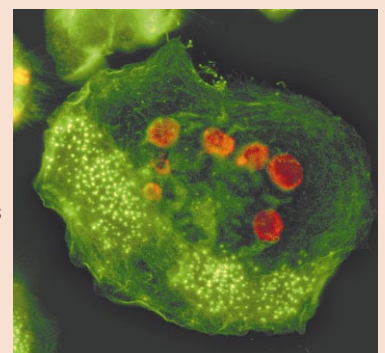
Although inflammatory responses are vital for host defence against infection, when persistent they also underlie important diseases such

as asthma or rheumatoid arthritis. Dangerous immune cells can be quietly cleared from inflamed sites by undergoing apoptosis followed by engulfment by phagocytes, promoting the resolution of acute inflammation². Moreover, the uptake of apoptotic cells actively suppresses the secretion from activated macrophages of pro-inflammatory mediators such as tumour necrosis factor- α ($\text{TNF-}\alpha$)^{26,27}. Safe clearance might therefore be doubly beneficial in inflammatory responses, preventing the secondary necrosis of apoptotic cells, with associated uncontrolled release of injurious contents, and 'calming' pro-inflammatory macrophages.

How is anti-inflammatory meaning conferred on the phagocytic clearance of apoptotic cells? 'Resetting' of activated macrophages can be mimicked by the ligation of macrophage receptors mediating the engulfment of apoptotic cells, notably CD36 (ref. 26), its 'bridging' ligand thrombospondin²⁶ and a newly discovered receptor for PtdSer exposed by apoptotic cells²⁸. Indeed, receptor-triggered release of the anti-inflammatory and immunosuppressive cytokine transforming growth factor- $\beta 1$ ($\text{TGF-}\beta 1$) by macrophages ingesting apoptotic cells might be crucial in mediating the autocrine or paracrine suppression of macrophage-directed inflammation²⁶⁻²⁸. Macrophages are nevertheless activated to secrete pro-inflammatory mediators by the ingestion of white blood cells undergoing secondary necrosis after apoptosis, but not by intact apoptotic cells²⁹. Therefore, whether clearance has an anti-inflammatory meaning might be determined by the state of the dying cell, the phagocyte receptors engaged and the downstream signalling pathways activated, which remain poorly understood.

Moreover, the anti-inflammatory action of phagocyte clearance of apoptotic cells might be perturbed in disease processes. For example, anti-phospholipid autoantibodies that recognize and bind PtdSer exposed by apoptotic cells can be found in patients with

Figure 2 A human monocyte-derived macrophage ingests multiple apoptotic bodies. Jurkat T-cell targets were labelled with 5-(and 6)-carboxytetramethylrhodamine succinimidyl ester and irradiated to induce apoptosis. The macrophages were stained with fluorescein isothiocyanate-conjugated phalloidin to identify actin filaments. (Photograph courtesy of M. Janes and P. Henson).



persistent and relapsing inflammatory disorders such as systemic lupus erythematosus (SLE). Such autoantibodies can coat apoptotic cells so that they are bound by macrophage Fc receptors with the result that TNF- α release is promoted rather than suppressed³⁰, threatening conversion of the anti-inflammatory clearance of dying cells into a pro-inflammatory event.

Control of macrophage-directed cell death

Macrophages are not merely scavengers of dying cells; they can also direct the death of unwanted cells during tissue remodelling. For example, elegant experiments on macrophage depletion and repletion have demonstrated a key role in the deletion of surplus blood vessels in the developing rodent eye³¹. Macrophages recruited from the blood can dock onto and direct apoptosis in unwanted microvascular endothelial cells. Furthermore, activated macrophages can also delete resident cells during the remodelling of inflamed sites in adult mammals, inducing apoptosis in neighbouring cells by mechanisms including nitric oxide release³².

The capacity for macrophages to direct cell death is regulated by the uptake of apoptotic cells. For example, the cytolysis of tumour cells by activated macrophages is inhibited by the ingestion of apoptotic but not necrotic cells³³. This might explain the surprising cohabitation of macrophages and malignant cells in many tumours. Blockade of the ingestion of cancer cells undergoing apoptosis by macrophages in the tumour might therefore be a new therapeutic approach in malignancy, disengaging an undesirable brake on macrophage cytotoxic capacity. Moreover, blocking the anti-cytotoxic response of phagocytes taking up apoptotic cells could provide a new approach towards the treatment of diseases caused by protozoans. These organisms can exploit the suppressive effects of the ingestion of apoptotic cells to evade macrophage-directed killing of parasites. For example, the protozoan parasite *Trypanosoma cruzi* causes the cardiac illness Chagas's disease, in which T cells die by activation-induced apoptosis. By mechanisms including the ligation of macrophage $\alpha_v\beta_3$, the local release of TGF- β_1 and prostaglandin E_2 (PGE $_2$), and the generation of polyamines necessary for parasite growth consequent on the induction of macrophage ornithine decarboxylase activity, the uptake of apoptotic T cells inhibits the generation of protective nitric oxide by macrophages and promotes the invasion of macrophages by *T. cruzi*³⁴. But inhibiting PGE $_2$ release from phagocytic macrophages with cyclo-oxygenase inhibitors such as aspirin or indomethacin not only protects macrophages from the induction of ornithine decarboxylase activity and infection by parasites *in vitro*, but also markedly decreases parasitaemia in mice infected with *T. cruzi*³⁴.

We therefore already know that inhibiting the suppressive effects of apoptotic cell ingestion after the macrophage-directed death of

'unwanted' cancer or parasite cells could have therapeutic use. Similar principles could be important in manipulating the remodelling of tissues injured by inflammatory and vascular diseases. But caution is required because there might be dual meaning in the ingestion of apoptotic cells by macrophages at sites of tissue remodelling. In some circumstances, far from suppressing macrophage cytotoxic capacity, the uptake of apoptotic cells can trigger the release of the death-inducing cytokine CD95 ligand (Apo-1/Fas ligand) by macrophages, resulting in the apoptosis of bystander cells³⁵.

Regulation of immune responses

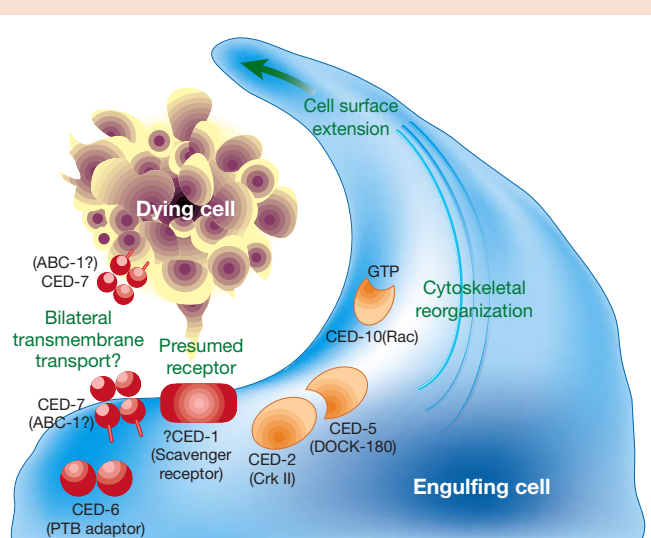
Glimpses of further consequences of the clearance of dying cells by phagocytes came from the observations that bone-marrow-derived dendritic cells could ingest apoptotic cells³⁶ and could present antigen derived from ingested apoptotic cells to T cells, thus initiating immune responses mediated by cytotoxic lymphocytes^{37,38}. Dendritic cells are specialized for the presentation of ingested antigen to lymphocytes but do so most efficiently if they subsequently receive a maturation or 'danger' signal such as an exposure to microorganisms³⁹. The phagocytosis of dying cells by dendritic cells is an important candidate mechanism for the promotion of immune responses to organisms capable of inducing apoptosis in host cells. Dendritic cells can also present antigen derived from ingested apoptotic cells to helper T cells³⁹. Exactly how peptides from ingested cells are presented on major histocompatibility complex (MHC) molecules by dendritic cells needs to be clarified, but dendritic cells use the $\alpha_v\beta_3$ integrin rather than $\alpha_v\beta_3$ in the engulfment of apoptotic cells, which could prove to be significant^{19,40}.

Nevertheless, the presentation of apoptotic-cell-derived antigens by dendritic cells must be subject to exquisite controls, as this threatens to incite immunity against self components⁴¹, particularly because protein cleavage within dying cells might generate 'neo-autoantigens'⁴². Indeed, the engulfment of apoptotic cells by dendritic cells can be observed in health. For example, migratory dendritic cells transport apoptotic intestinal epithelial cells to T-cell areas of mesenteric lymph nodes in apparently healthy rats⁴³. Circumstantial evidence^{39,44} suggests that in the absence of maturation signals, the uptake by dendritic cells of self components packaged in apoptotic cells might reinforce tolerance to self tissue so that when tissue injury does occur, T cells likely to respond to self antigens have been anergized or deleted. This fascinating idea needs to be tested, but the immune system seems able to detect two classes of meaning in the phagocytosis of apoptotic cells by dendritic cells: a message to respond to non-self coupled with a reminder not to attack self³⁹.

Available data indicate that a 'pro-immune meaning' can be conferred on the phagocytosis of dying cells by dendritic cells when

Figure 3 Signalling the engulfment of dying cells in *Caenorhabditis elegans*.

Mutations in six genes are known to affect the engulfment of cell corpses by non-professional neighbouring cells in this nematode. CED-2, CED-5 and CED-10 intracellular proteins signal in a manner comparable to their respective mammalian homologues CrkII, DOCK 180 and Rac, mediating the cytoskeletal reorganization and extension of the engulfing cell surface around the dying cell. CED-7, homologous with mammalian phagocyte ABC-1, acts in both dying and engulfing cells⁶, possibly in transmembrane lipid transport. We speculate that CED-1, yet to be characterized, is analogous to mammalian scavenger receptors; CED-7 and CED-1 probably promote engulfment by interacting with the signalling adaptor protein CED-6. PTB, phosphotyrosine binding.



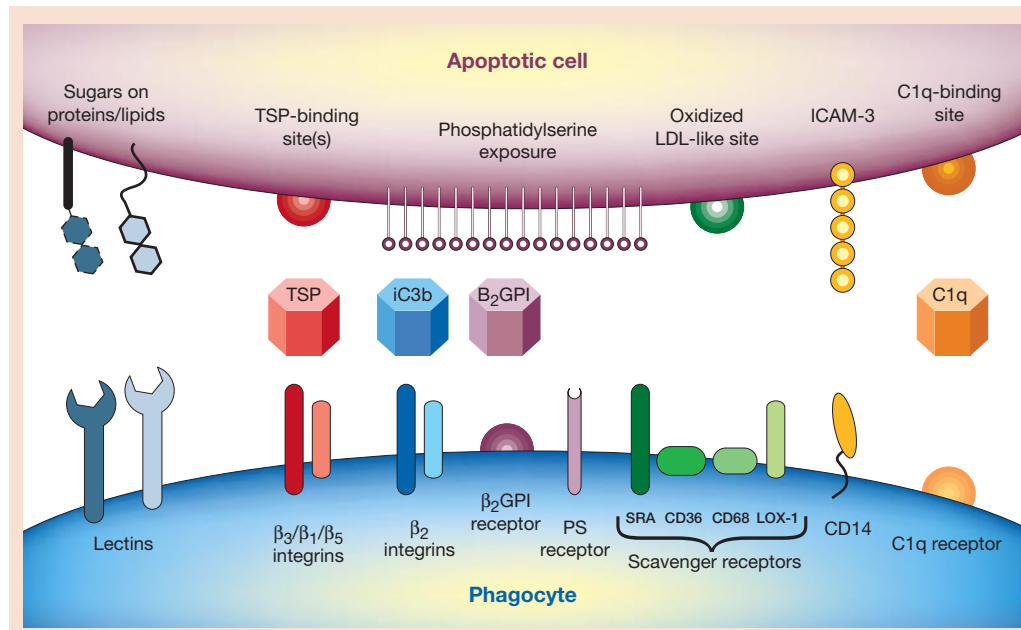


Figure 4 The phagocyte recognition array in the mammalian clearance of apoptotic cells. Inhibitor studies in assays of the ingestion of apoptotic cells *in vitro* by phagocytes have revealed a variety of candidate molecules, many of which have incompletely understood roles or uncharacterized binding partners, as discussed in the text. A repertoire of 'eat me' signals (top) interact with receptors on the phagocyte (bottom), either directly or via serum-derived bridging molecules (middle). LDL, low-density lipoprotein; SRA, class A scavenger receptor; TSP, thrombospondin.

there is defective clearance of apoptotic cells. For example, humans deficient in C1q almost invariably develop SLE. C1q-knockout mice are also highly susceptible to the development of a SLE-like disorder¹⁰. Inflamed tissues in these mice contain many free cells undergoing death by apoptosis, consistent with demonstrable defects in the clearance of apoptotic cells by macrophages¹¹. Therefore, in C1q-deficient individuals, secondary necrosis of apoptotic cells escaping clearance could provide 'danger' signals that might switch the presentation by dendritic cells of self peptides derived from ingested apoptotic cells from 'tolerogenic' to 'pro-immune'^{39,45}. This might explain why SLE is characterized by autoimmunity directed against antigens that normally reside within cells. Agents that promote the safe clearance of dying cells by phagocytes, such as glucocorticoids⁴⁶ and lipoxins⁴⁷, could be exploited in therapies.

Deciphering lipid-coded clearance

Decoding still further the meaning of the clearance of dying cells by phagocytes will provide new insights into the pathogenesis and treatment of inflammatory, autoimmune, malignant and infective conditions. The recent characterization of a complete system for PtdSer-directed phagocytosis of apoptotic cells indicates that surface lipids on dying cells might determine the meaning of clearance. Exposure of PtdSer, mediated by the poorly understood inhibition of an aminophospholipid translocase and the activation of a lipid scramblase (which is itself dependent on the entry of Ca^{2+} ions and the phosphorylation of scramblase), reveals an 'eat me' flag that could direct a range of phagocyte responses^{9,48,49}. A newly identified stereospecific PtdSer receptor expressed by professional and amateur phagocytes seems to be coupled to anti-inflammatory clearance, stimulating macrophage TGF- β 1 secretion and inhibiting lipopolysaccharide-induced TNF- α release²⁸. However, although other receptors have broader lipid specificity, exposure to PtdSer might also promote the direct ligation of apoptotic cells by CD36 (ref. 50), CD68 (ref. 14), CD14 (ref. 18) and LOX-1 (ref. 15) on the phagocyte surface, with as yet unknown consequences. Further complexity is added by the possible bridging of PtdSer to other phagocyte receptors by soluble proteins such as $\beta_2\text{GPI}$ ¹³ and opsonic complement fragments such as iC3b¹². Therefore, even the single cipher of PtdSer exposure could encode a range of meanings into clearance of dying cells.

Conclusions and further challenges

The clearance of apoptotic cells by phagocytes initially attracted interest as the final act in programmed cell death and an intriguing

problem in the discrimination of self from unwanted self. The genetics of *C. elegans* has provided a basic understanding of the signalling of engulfment, and studies of mutations affecting the clearance of dying cells by professional phagocytes in *Drosophila* will help to unravel the complexity inferred from inhibitor studies in mammalian systems. However, the recent revelation that the clearance of apoptotic cells by phagocytes defines meaning for cell death indicates that studies in mammals *in vivo* are urgently required to address regulatory roles in inflammation, immune responses and tissue remodelling. Dead men may tell no tales, but dead cells certainly do, the phagocyte having the last word. □

- Kerr, J. F. R., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with widespread implications in tissue kinetics. *Br. J. Cancer* **26**, 239–257 (1972).
- Savill, J. Apoptosis: phagocytic docking without shocking. *Nature* **392**, 442–443 (1998).
- Platt, N., da Silva, R. P. & Gordon, S. Recognising death: the phagocytosis of apoptotic cells. *Trends Cell Biol.* **8**, 365–372 (1998).
- Wu, Y. C. & Horvitz, H. R. *C. elegans* phagocytosis and cell-migration protein CED-5 is similar to human DOK180. *Nature* **392**, 501–504 (1998).
- Lui, Q. A. & Hengartner, M. O. Candidate adaptor protein CED-6 promotes the engulfment of apoptotic cells in *C. elegans*. *Cell* **93**, 961–972 (1998).
- Wu, Y. C. & Horvitz, H. R. The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to ABC transporters. *Cell* **93**, 951–960 (1998).
- Reddien, P. W. & Horvitz, H. R. CED-2/Crkl and CED-10/Rac control phagocytosis and cell migration in *Caenorhabditis elegans*. *Nature Cell Biol.* **2**, 131–135 (2000).
- Fadok, V. A., Bratton, D. L., Frasch, S. C., Warner, M. L. & Henson, P. M. The role of phosphatidylserine in recognition of apoptotic cells by phagocytes. *Cell Death Differ.* **5**, 557–563 (1998).
- Dini, L., Carla, E. C., Deluca, M., Faraldi, G. & Tagliaferro, G. Phagocytosis of apoptotic cells: liver recognition and molecular mechanisms. *Ital. J. Zool.* **66**, 317–322 (1999).
- Botto, M. *et al.* Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies. *Nature Genet.* **19**, 56–59 (1998).
- Taylor, P. R. *et al.* A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells. *J. Exp. Med.* **192**, 359–366 (2000).
- Mevorach, D., Mascarenhas, J. O., Gershov, D. & Elkon, K. B. Complement-dependent clearance of apoptotic cells by human macrophages. *J. Exp. Med.* **188**, 2313–2320 (1998).
- Balasubramanian, K., Chandra, J. & Schroit, A. J. Immune clearance of phosphatidylserine-expressing cells by phagocytes. The role of β_2 -glycoprotein I in macrophage recognition. *J. Biol. Chem.* **272**, 31113–31117 (1997).
- Sambrano, G. R. & Steinberg, D. Recognition of oxidatively damaged and apoptotic cells by an oxidized low density lipoprotein receptor on mouse peritoneal macrophages: role of membrane phosphatidylserine. *Proc. Natl Acad. Sci. USA* **92**, 1396–1400 (1995).
- Oka, K. *et al.* Lectin-like oxidized low-density lipoprotein receptor 1 mediates phagocytosis of aged/apoptotic cells in endothelial cells. *Proc. Natl Acad. Sci. USA* **95**, 9535–9540 (1998).
- Chang, M.-K. *et al.* Monoclonal antibodies against oxidized low-density lipoprotein bind to apoptotic cells and inhibit their phagocytosis by elicited macrophages: evidence that oxidation-specific epitopes mediate macrophage recognition. *Proc. Natl Acad. Sci. USA* **96**, 6353–6358 (1999).
- Moffatt, O. D., Devitt, A., Bell, E. D., Simmons, D. L. & Gregory, C. D. Macrophage recognition of ICAM-3 on apoptotic leukocytes. *J. Immunol.* **162**, 6800–6810 (1999).
- Devitt, A. *et al.* Human CD14 mediates recognition and phagocytosis of apoptotic cells. *Nature* **392**, 505–509 (1998).
- Albert, M. L., Kim, J.-I. & Birge, R. B. The $\alpha_5\beta_1$ integrin recruits the CrkII/Dock180/Rac1 molecular complex for phagocytosis of apoptotic cells. *Nature Cell Biol.* (in press).

20. Smits, E., Crieckinge, W. V., Plaetnick, G. & Bogaert, T. The human homologue of *Caenorhabditis elegans* CED-6 specifically promotes phagocytosis of apoptotic cells. *Curr. Biol.* **9**, 1351–1354 (1999).
21. Marguet, D., Luciani, M. F., Moynault A., Williamson, P. & Chimini, G. Engulfment of apoptotic cells involves the redistribution of membrane phosphatidylserine on phagocyte and prey. *Nature Cell Biol.* **1**, 454–456 (1999).
22. Hamon, Y. *et al.* ABC1 promotes engulfment of apoptotic cells and transbilayer redistribution of phosphatidylserine. *Nature Cell Biol.* **2**, 399–406 (2000).
23. Franc, N. C., Heitzler, P., Ezekowitz, A. B. & White, K. Requirement for Croquemort in phagocytosis for apoptotic cells in *Drosophila*. *Science* **284**, 1991–1994 (1999).
24. Platt, N., Suzuki, H., Kodama, T. & Gordon, S. Apoptotic thymocyte clearance in scavenger receptor class A-deficient mice is apparently normal. *J. Immunol.* **164**, 4861–4867 (2000).
25. Meagher, L. C., Savill, J. S., Baker, A. & Haslett, C. Phagocytosis of apoptotic neutrophils does not induce macrophage release of thromboxane B₂. *J. Leuk. Biol.* **52**, 269–273 (1992).
26. Voll, R. E., Herrmann, M., Roth, E. A., Stach, C. & Kalden, J. R. Immunosuppressive effects of apoptotic cells. *Nature* **390**, 350–351 (1997).
27. Fadok, V. A. *et al.* Macrophages that have ingested apoptotic cells *in vitro* inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF- β , PGE₂, and PAF. *J. Clin. Invest.* **101**, 890–898 (1998).
28. Fadok, V. A. A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature* **405**, 85–90 (2000).
29. Stern, M., Savill, J. & Haslett, C. Human monocyte-derived macrophage phagocytosis of senescent eosinophils undergoing apoptosis: mediation by $\alpha\beta_2$ /CD36 thrombospondin recognition mechanism and lack of phlogistic response. *Am. J. Pathol.* **149**, 911–921 (1996).
30. Manfredi, A. A. *et al.* Apoptotic cell clearance in systemic lupus erythematosus. I. Opsonization by antiphospholipid antibodies. *Arth. Rheum.* **41**, 205–214 (1998).
31. DiezRouz, G. & Lang, R. A. Macrophages induce apoptosis in normal cells *in vivo*. *Development* **124**, 3633–3638 (1997).
32. Duffield, J. A. *et al.* Activated macrophages direct apoptosis and suppress mitosis of mesangial cells. *J. Immunol.* **164**, 2110–2119 (2000).
33. Reiter, I., Krammer, B. & Schwamberger, G. Differential effect of apoptotic versus necrotic tumor cells on macrophage antitumor activities. *J. Immunol.* **163**, 1730–1732 (1999).
34. Freire-de-Lima, C. G. *et al.* Uptake of apoptotic cells drives the growth of a pathogenic trypanosome in macrophages. *Nature* **403**, 199–203 (2000).
35. Brown, S. B. & Savill, J. Phagocytosis triggers macrophage release of Fas ligand and induces apoptosis of bystander leukocytes. *J. Immunol.* **162**, 480–485 (1999).
36. Rubartelli, A., Foggi, A. & Zocchi, M. K. The selective engulfment of apoptotic bodies by dendritic cells is mediated by the $\alpha\beta_2$ integrin and requires intracellular and extracellular calcium. *Eur. J. Immunol.* **27**, 1893–1900 (1997).
37. Bellone, M. *et al.* Processing of engulfed apoptotic bodies yields T cell epitopes. *J. Immunol.* **159**, 5391–5399 (1997).
38. Albert, M. L., Sauter, B. & Bhardwaj, N. Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs. *Nature* **392**, 86–89 (1998).
39. Steinman, R. M., Turley, S., Mellman, I. & Inaba, K. The induction of tolerance by dendritic cells that have captured apoptotic cells. *J. Exp. Med.* **191**, 411–416 (2000).
40. Albert, M. L. *et al.* Immature dendritic cells phagocytose apoptotic cells via $\alpha\beta_2$ and CD36, and cross-present antigens to cytotoxic T lymphocytes. *J. Exp. Med.* **188**, 1359–1368 (1998).
41. Mevorach, D., Zhou, J. L., Song, X. & Elkon, K. B. Systemic exposure to irradiated apoptotic cells induces autoantibody production. *J. Exp. Med.* **188**, 387–392 (1998).
42. Casciola-Rosen, L. A., Anhalt, G. J. & Rosen, A. DNA-dependent protein kinase is one of a subset of autoantigens specifically cleaved early during apoptosis. *J. Exp. Med.* **182**, 1625–1634 (1995).
43. Huang, F.-P. *et al.* A discrete subpopulation of dendritic cells transports apoptotic intestinal epithelial cells to T cell areas of mesenteric lymph nodes. *J. Exp. Med.* **191**, 435–443 (2000).
44. Nakamura, K. *et al.* Unresponsiveness of peripheral T cells induced by apoptotic bodies derived from autologous T cells. *Cell. Immunol.* **193**, 147–154 (1999).
45. Rovere, P. *et al.* Bystander apoptosis triggers dendritic cell maturation and antigen-presenting function. *J. Immunol.* **161**, 4467–4471 (1998).
46. Liu, Y. *et al.* Glucocorticoids promote nonphlogistic phagocytosis of apoptotic leukocytes. *J. Immunol.* **162**, 3639–3646 (1999).
47. Godson, C. *et al.* Lipoxins rapidly stimulate nonphlogistic phagocytosis of apoptotic neutrophils by monocyte-derived macrophages. *J. Immunol.* **164**, 1663–1667 (2000).
48. Frasch, S. C. *et al.* Regulation of phospholipid scramblase activity during apoptosis and cell activation by protein kinase C δ . *J. Biol. Chem.* **275**, 23065–23073 (2000).
49. Wiedmer, T., Zhou, Q., Kwok, D. Y. & Sims, P. J. Identification of three new members of the phospholipid scramblase gene family. *Biochim. Biophys. Acta* **1467**, 244–253 (2000).
50. Ren, Y., Silverstein, R. L., Allen, J. & Savill, J. CD36 gene transfer confers capacity for phagocytosis of cells undergoing apoptosis. *J. Exp. Med.* **181**, 1857–1862 (1995).

Acknowledgements

We thank the many scientists working in the field, especially those whose work could not be cited because of space constraints; we also thank C. Gilchrist for secretarial assistance. Support from the Wellcome Trust, the Medical Research Council, the National Kidney Research Fund and the National Institutes of Health is gratefully acknowledged.

CD95's deadly mission in the immune system

Peter H. Krammer

Tumorimmunology Program, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

Apoptosis in the immune system is a fundamental process regulating lymphocyte maturation, receptor repertoire selection and homeostasis. Thus, death by apoptosis is as essential for the function of lymphocytes as growth and differentiation. This article focuses on death receptor-associated apoptosis and the role of CD95 (Apo-1/Fas)-mediated signalling in T-cell and B-cell development and during the course of an immune response. Gaining an insight into these processes improves our understanding of the pathogenesis of diseases such as cancer, autoimmunity and AIDS, and opens new approaches to rational treatment strategies.

The immune system is a society of interacting cells consisting of T and B lymphocytes, natural killer (NK) cells, macrophages and professional antigen-presenting cells (APCs) and their various subclasses. Most cellular components of the immune system are born in the bone marrow. B lymphocytes, NK cells and macrophages mature in the bone marrow and in the fetal liver. T lymphocytes mature in the bone marrow and in the thymus. T-cell and B-cell development share many features, yet differ in others (Fig. 1).

B cells

B cells express cell membrane receptors (antibodies) with a single antigen specificity. Millions of different B cells produce antibodies and can potentially capture millions of antigens. The sum of all antibody specificities is called 'the antibody repertoire'¹. B cells are selected in the bone marrow on the basis of the affinity of their antibodies: cells with high affinity for proteins derived from 'self' tissues are eliminated. Mature B lymphocytes leave the bone marrow and populate the secondary lymphoid organs, spleen and lymph nodes and the gut-associated lymphoid tissue. Once activated by an antigen, B cells undergo a second round of selection in the follicles of secondary lymphoid organs, after which they mature into plasma cells that produce and secrete antigen-specific antibodies, and then recirculate to the bone marrow².

T cells

Pre-T lymphocytes emigrate from the bone marrow into the thymus. In the thymus, they mature and are positively or negatively selected, depending on the affinity of their T-cell antigen receptors (TCRs) for self major histocompatibility (MHC) antigens. MHC class I and II antigens are molecules that sample peptide fragments from foreign and self proteins, respectively. They display these peptides at the cell's surface for scrutiny by T cells — a process called 'antigen presentation'. Each MHC class I or II protein presents a different fragment; thousands of MHC molecules protrude from every cell. Most of the peptides presented in the thymus are derived from self proteins. T cells with a high affinity for self MHC molecules and peptide are eliminated to ensure tolerance to normal tissues and to prevent autoimmunity. T cells that interact with MHC class II molecules develop into cells that express the CD4 molecule on their surface (CD4⁺), and those with affinity for MHC class I molecules turn into T lymphocytes that carry the CD8 antigen (CD8⁺). Only

mature T cells that produce a functional TCR leave the thymus and populate the secondary lymphoid organs. Mature CD4⁺ T cells function as helper T cells and secrete cytokines that regulate either cellular immune responses (T helper 1 cells) or antibody responses (T helper 2 cells). Mature CD8⁺ T cells function as cytotoxic effector (killer) cells³ (Fig. 2).

Cells of the immune system work as a team. After an attack by infectious agents, professional APCs, dendritic cells in particular, present antigenic peptides of these infectious agents to T cells⁴. Antigen–MHC complexes, co-stimulating cell-surface molecules on APC and cytokines drive T cells into clonal expansion. These T cells, in turn, then communicate with other T or B cells to regulate their responses. After a peak phase with the highest clonal expansion of reactive cells, the immune response undergoes a down phase. Most lymphocytes are eliminated; the few that survive constitute the pool of memory cells⁵.

Life and death in the immune system

Several features of the immune system are unique. One is its specificity: the repertoire of T and B lymphocytes, initially built from randomly selected antibody and TCR variable-region genes, is shaped by selection to cope, on the one hand, with the vast universe of antigens and, on the other hand, with the danger of autoimmunity³. Another distinctive feature is its homeostatic control: after a clonal expansion phase, antigen-reactive lymphocytes must be titrated back until the pool of lymphoid cells reaches the baseline level again⁶. This is achieved by balanced fine-tuning between growth/expansion and death by apoptosis; generally, the immune system produces more cells than finally needed, and extra cells are eliminated by apoptosis.

Apoptosis is the most common form of death in cells of the immune system. It is astounding how many different pathways immune cells can choose from to die. In principle, death can be by neglect when the antigen-specific receptors of the lymphoid cells are not stimulated or the lymphocytes are deprived of trophic cytokines. In a more active form, death can involve the death-receptor/death-ligand systems^{7–9}. Apoptosis is such a central regulatory feature of the immune system that it is not surprising that too little or too much apoptosis results in severe diseases.

The CD95–CD95L death system

A subset of tumour necrosis factor receptor (TNF-R) family members is involved in death transducing signals and are

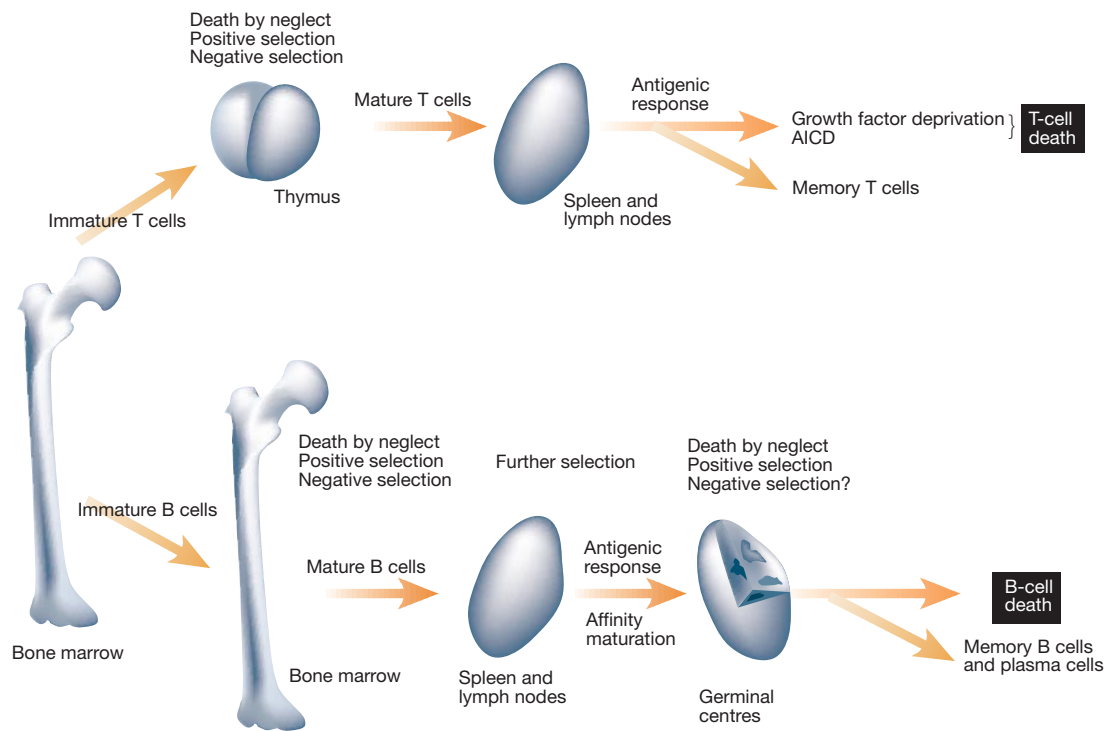


Figure 1 T- and B-lymphocyte development (see text). AICD, activation-induced cell death.

therefore referred to as the 'death receptors.' Members of this family contain one to five cysteine-rich repeats in their extracellular domain, and a death domain in their cytoplasmic tail. The death domain is essential for transduction of the apoptotic signal. CD95 is one such family member and has a significant role in the immune system and outside. It is a widely expressed glycosylated cell-surface molecule of relative molecular mass ~45,000–52,000 (335 amino-acid residues). It is a type I transmembrane receptor, but alternative splicing can result in a soluble form, the function of which is unclear¹⁰. CD95 expression can be boosted by cytokines such as interferon- γ and TNF but also by the activation of lymphocytes^{11,12}. CD95-mediated apoptosis is triggered by its natural ligand, CD95L, which is a TNF-related type II transmembrane molecule⁷ and expressed in a far more restricted way than the receptor. Killer cells (so-called cytotoxic T lymphocytes) remove, for example, virus-infected cells, and those that express CD95L can do so by interacting with the CD95 receptor^{13,14} on their targets (Fig. 2). CD95L is seen on killer cell-derived vesicles^{15,16}, but can also be cleaved from the membrane by a metalloprotease^{17–19}; whereas soluble human CD95L can induce apoptosis²⁰, soluble mouse CD95L cannot²¹.

The CD95 death-inducing signalling complex

The oligomerization, most probably the trimerization, of CD95 is required for transduction of the apoptotic signal. A complex of proteins associates with activated CD95 (refs 8, 22). This death-inducing signalling complex (DISC) forms within seconds of receptor engagement. First, the adaptor FADD (Fas-associated death domain protein, also known as Mort1) binds via its own death domain to the death domain in CD95. FADD also carries a so-called death-effector domain (DED), and, again by homologous interaction, recruits the DED-containing procaspase-8 (also known as FLICE) into the DISC. Next, procaspase-8 is activated proteolytically and active caspase-8 is released from the DISC into the cytoplasm in the form of a heterotrimer of two small subunits and two large subunits²³. Active caspase-8 cleaves various proteins in the cell including procaspase-3, which results in its activation and the completion of the cell death

programme. Various other proteins have been described to bind to activated CD95 and the DISC, but their precise role and importance in the regulation of apoptosis remain to be defined.

Recently, with the use of fluorescence resonance energy transfer, another model of CD95 signalling has been worked out. Extracellular pre-ligand-binding assembly domains (PLADs) were described for CD95 and TNF-R, which are supposed to aggregate the receptors before ligand binding. To prevent the premature signalling of pre-associated receptors, which is a dangerous situation, intracellular receptor-associated apoptosis blockers were postulated^{24,25}. On the basis of the PLAD model it is not entirely clear how ligand binding interferes with PLAD association and leads to receptor association, which initiates apoptosis. More structural work is needed to resolve these issues. It is also unclear whether the DISC model and the PLAD model complement each other to describe initial signalling events *in vivo*.

Whereas some cytotoxic T lymphocytes kill their target cells by turning on the death receptor, others use perforin and granzyme B (GrB) to eliminate infected cells. With the help of perforin, GrB finds its way into the target cell and can kill it by directly cleaving and activating caspase-8 (ref. 14) (Fig. 2).

Two different pathways downstream of CD95

In so-called type I cells²⁶, the death signal is propagated by a caspase cascade initiated by the activation of large amounts of caspase-8 at the DISC, followed by a rapid cleavage of caspase-3 and other caspases, which in turn cleave vital substrates in the cell. But in type II cells²⁶ hardly any DISC is formed, so the caspase cascade cannot be propagated directly but has to be amplified via the mitochondria. Caspase-8 cuts the Bcl-2 family member Bid; truncated Bid 'activates' the mitochondria^{27,28}. Mitochondria are 'activated' in both type I and type II cells, but are not strictly necessary for the death of type I cells. On 'activation', mitochondria release pro-apoptotic molecules such as cytochrome *c*²⁹ and Smac/DIABLO^{30,31}. Together with the apoptosis protease-activating factor Apaf-1 and procaspase-9 in the cytoplasm, these molecules form the so-called apoptosome, the second initiator complex of apoptosis (see review in this issue by Hengartner,

pages 770–776). Caspase-9 activates further downstream caspases and the end result, again, is apoptosis (Fig. 3). The reason for the differences between type I and type II cells is unclear at present, but perhaps biochemical differences at the receptor level hold the answer. Recently, Bid-deficient mice were generated³² that are resistant to CD95-induced hepatocellular apoptosis. In Bid^{-/-} cells from these mice, mitochondrial dysfunction was delayed, cytochrome *c* was not released, effector caspase activity was reduced and the cleavage of apoptosis substrates was altered. Taken together, these data support the type-I/type-II concept of CD95 signalling.

FLIPs (FLICE-inhibitory proteins)

Additional DED-containing proteins have been found in a certain class of herpes viruses. These proteins contain two DEDs and also bind to the CD95–FADD complex. This inhibits the recruitment and activation of caspase-8, formerly known as FLICE; hence their name FLIPs (for FLICE-inhibitory proteins). In transfected cells, v-FLIP inhibited the apoptosis induced by several apoptosis-inducing receptors (CD95, TNF-R1, TRAMP/DR3 and TRAIL-R1), indicating that these receptors use similar signalling pathways^{33–35}. Two human homologues of v-FLIP have been identified by several groups at the same time and are known under a variety of names. Although it is widely assumed that the cellular FLIPs block apoptosis as well, the data are ambiguous and c-FLIP might be pro-apoptotic or anti-apoptotic depending on the cellular context. Recent results with cells from c-FLIP-deficient mice support the role of c-FLIP as an anti-apoptotic molecule³⁶.

Signalling by other death receptors

Signalling of apoptosis by other members of the death receptor subfamily seems to follow the same basic rules and is initiated by the same sequential steps^{37–39}: (1) ligand binding, (2) receptor trimerization and DISC formation, (3) attraction of the adapter molecule FADD into the DISC, (4) association of procaspase-8, and (5) auto-catalytic cleavage of the procaspase and the formation of active caspase-8 (a heterotetramer). Active caspase-8 then serves as the initiator caspase, activating other further downstream executioner caspases that cleave cellular death substrates, resulting in the morphological and biochemical catastrophe termed apoptosis. The following paragraphs discuss briefly how these signalling pathways are used in the immune system.

T-lymphocyte death in the thymus

The T-cell repertoire is shaped in the thymus by apoptosis and survival signals. A young adult mouse with $(1–2) \times 10^8$ thymocytes generates between 20 and 40 million new T cells per day⁴⁰. But the number of T cells that leave the thymus and enter the peripheral T-cell pool is only about 2–3% of the number initially generated.

Despite the high death rate of T cells in the thymus, only a limited number of apoptotic cells can be observed in histological sections. Thus, apoptotic thymocytes are removed efficiently and, most importantly, this is achieved without signs of inflammation (see refs 41–43 and the review in this issue by Savill and Fadok, pages 784–788).

Pre-T lymphocytes, after entry into the thymus, differentiate and rearrange their TCR genes. Those T cells that fail to rearrange their TCR genes productively and thus cannot be stimulated by self-MHC–peptide complexes die by neglect. In T lymphocytes of FADD dominant-negative transgenic mice the requirement for pre-TCR signals is bypassed⁴⁴. In these mice, T-cell survival and differentiation are promoted. Because FADD is an essential adaptor of several death-receptor DISCs, these data suggest a role for death receptors at this early stage of T-cell development. Thymocytes that have successfully passed pre-TCR selection mature further, develop into CD4⁺CD8⁺ (double-positive) T cells and undergo further TCR-affinity-driven positive and negative selection on thymic stromal cells. After these selection processes, mature single-positive CD4⁺ MHC class-II-restricted and CD8⁺ MHC class-I-restricted T cells leave the thymus and generate the peripheral T-cell pool. Like crossing several borders, the T cell crosses several checkpoints to ensure self-MHC restriction and self-tolerance.

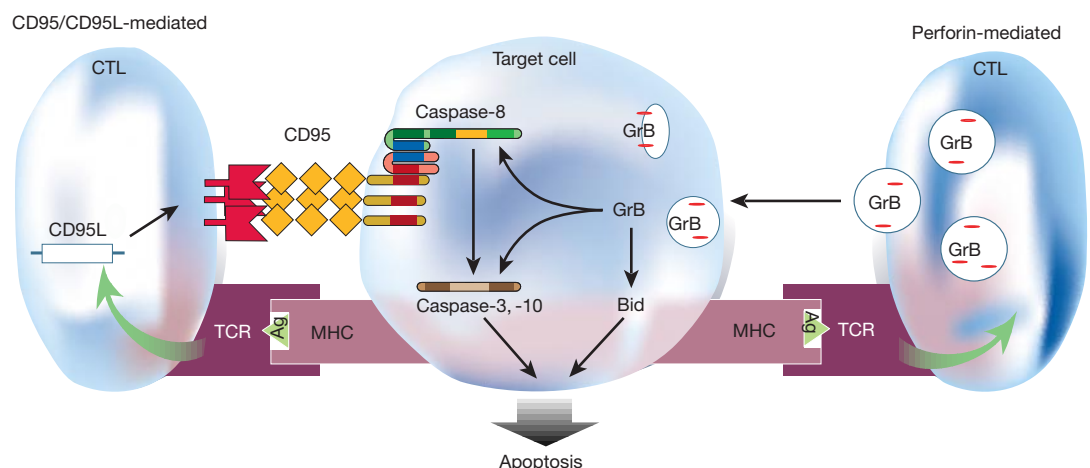
Initially, most investigators agreed that the CD95 system is not involved in negative selection because the TCR repertoire in mice with a defect in this system (*lpr*, *lpr^g* and *gld* mice) was not altered⁴⁵. But on closer inspection it was found that negative selection might involve the CD95 system when T cells encounter high antigen concentrations⁴⁶. The role of other members of the TNF-R superfamily, TNF-R1 and TNF-R2, CD30 and CD40, remains controversial. Similarly, survival signals of thymocytes at different maturation stages remain ill-defined. Numerous data suggest that members of the Bcl-2 family influence survival of immature T lymphocytes, that is, positive selection, but not negative selection³.

Finally, a modulating role in thymocyte survival and apoptosis has been ascribed to several different molecules such as glucocorticoid hormones, cytokines, co-stimulating cell-surface receptors, signalling molecules, transcription factors³ and nitric oxide⁴⁷. In view of the available data, our understanding of the molecular basis of apoptosis and selection of T lymphocytes in the thymus still remains fragmentary.

Deletion of peripheral T cells by apoptosis

Deletion by apoptosis is also observed in mature peripheral T cells. It occurs by neglect in those T cells that are not sufficiently stimulated by growth signals⁴⁸ and, importantly, it occurs at the peak or the down phase of the immune response (Fig. 4)^{49–54} to downregulate the number of reactive cells and to terminate the immune response. This

Figure 2 Cytotoxic T lymphocytes (CTLs) can kill target cells by the CD95 (yellow)/CD95L (red) system (left) or by the perforin/granzyme B (GrB) system (right). CD95 signalling involves a caspase cascade (see text). The entry of GrB into target cells and its release involve perforin. GrB cleaves and activates caspases in the target cell. The target cell dies by apoptosis through either a CD95 or a GrB signalling event.



so-called activation-induced cell death (AICD) might also serve as a second line of defence against autoimmunity by deleting autoreactive cells in the periphery.

After activation, T cells go through several phases (Fig. 4): (1) an interleukin (IL)-2-dependent clonal expansion and effector phase after challenge with antigen, (2) a down phase in which most antigen-specific T cells are eventually eliminated, and (3) a phase in which certain T cells that survive the down phase enter the memory T-cell pool. In the first phase, T cells are resistant to apoptosis; memory cells are also thought to be relatively resistant to apoptosis⁵⁵. But in the down phase T cells become progressively more sensitive to apoptosis in the presence of IL-2. Thus, IL-2 serves a dual role: it is initially mandatory for clonal expansion and later for sensitizing T cells towards apoptosis⁵⁶. *In vitro*, T-cell activation leads to the expression of CD95L and permits T cells to eliminate neighbouring CD95-positive cells; those cells that secrete CD95L can commit suicide^{11,54}. But whereas the CD95–CD95L system seems to be used at the initiation of AICD, the TNF-R2/TNF- α system is important at a later phase^{57,58}. Furthermore, reactive oxygen species and perforin/GrB are also involved in AICD^{59,60} and can also occur in a caspase-independent manner⁶¹. Apparently, activated T cells have various ways of dying. The molecular basis for their choice and the identity of the pathways used *in vivo* remain to be determined.

Sensitive T cells behave like type I cells, form a CD95–DISC complex and initiate a caspase cascade that results in apoptosis. But resistant T cells have decreased DISC formation, and amplification via the mitochondria is blocked by upregulation of the anti-apoptotic molecule Bcl-x_L. The role of c-FLIP in this is controversial⁶². Taken together, T cells can die by different routes and the shift from apoptosis resistance to sensitivity coincides with a shift from a type II to a type I pathway of apoptosis.

Co-stimulation and T-cell survival

T cells activated by one signal by means of the TCR can be saved from AICD by a second signal from co-stimulatory molecules, adhesion molecules or cytokine receptors. CD28 is a major co-stimulating co-receptor expressed on T cells; it is stimulated by CD80 and CD86 expressed on APCs and functions to increase cytokine production and cytokine receptor induction. Under certain conditions CD28 can sensitize T cells towards apoptosis, but generally CD28 enhances the cell proliferation and viability of T cells.

The effect of co-stimulation was observed on three levels: (1) a strong upregulation of c-FLIP_s (refs 63, 64), (2) the upregulation of Bcl-x_L (ref. 65), and (3) the downregulation of CD95L messenger RNA and protein at a defined time (8–12 h after stimulation). Thus, co-stimulation blocks both the type I and type II pathways in T cells and, at least temporarily, also blocks CD95L expression⁶³. At present it is not clear how antigen-activated T cells with an apoptosis-resistant phenotype are turned into memory T cells, and whether memory T cells are locked in the apoptosis-resistant state, but molecules such as c-FLIP and Bcl-x_L might hold some of the answers.

B-lymphocyte death

Three cell-surface molecules are key elements in the regulation of B-cell life and death: B-cell receptor (BCR), CD40 and CD95. The stage of maturation and activation of the B cell, the quantity and quality of the signal provided, and the context of cytokines and other components of the cellular environment are all key factors in whether triggering the BCR, for example by antigens, induces survival or death^{66,67}. Evidence from studies of normal and malignant B cells suggests that BCR activation induces apoptosis by the mitochondrial pathway. But many components of the signalling pathway are still elusive. It is therefore unclear which signals link BCR stimulation to mitochondrial activation^{68,69}.

As in T cells co-stimulated by CD28, BCR-activated B cells can be rescued from apoptosis by co-stimulation by way of CD40 that has been activated by CD40L expressed on T cells and macrophages. This

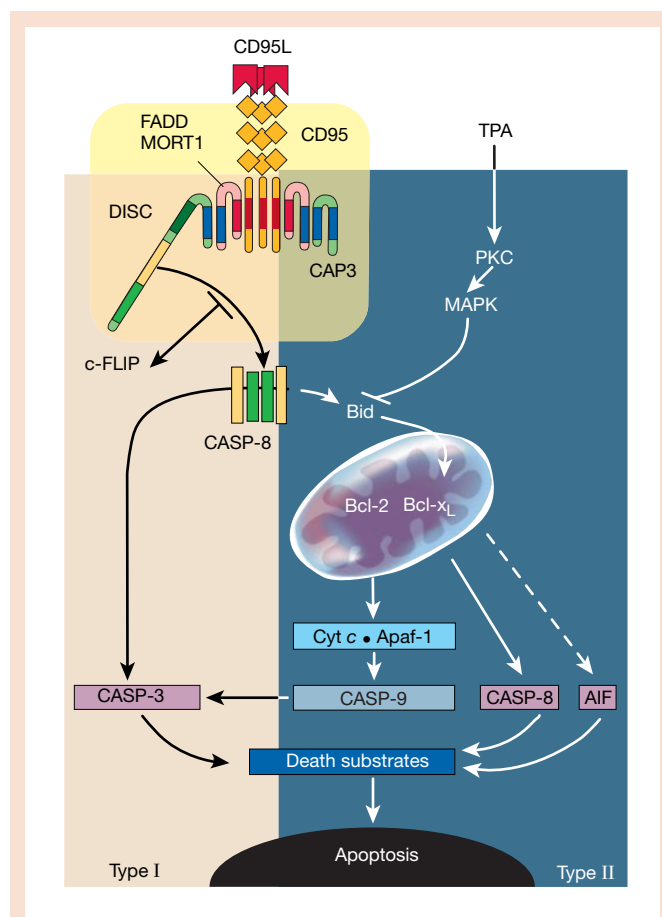


Figure 3 Signalling pathways induced by CD95. CD95 signalling pathway (including DISC formation) used in type I and type II cells (see text). TPA, 12-*O*-tetradecanoylphorbol-13-acetate; PKC, protein kinase C; MAPK, mitogen-activated protein kinase; CASP, caspase; AIF, apoptosis-initiating factor; CAP3, cytotoxicity-dependent Apo-1-associated protein 3).

stimulus might represent the most important survival signal for B cells even though such signals at a different maturation stage might also prepare B cells for death^{2,70}. Although it has been noted that transgenic *bcl-2* prevents death and impairs affinity maturation in germinal centres, it is unclear, for example, in which other situations Bcl-2 and other family members and the inhibitor-of-apoptosis (IAP) proteins block apoptosis⁷¹, and in which situations IL-4 and other cytokines act as survival signals². In addition, it is unclear how plasma cells die and which anti-apoptotic signals regulate their survival^{2,72}.

Thus, the principles of B-cell and T-cell development, repertoire selection and involvement of apoptosis in death by neglect and negative selection are similar. However, there are some fundamental B-cell-specific characteristics (Fig. 1).

Autoreactive B cells are eliminated in the bone marrow but, in response to antigenic stimulation, B cells undergo a second diversification and affinity maturation step in the germinal centres of the secondary lymphoid organs by a process called somatic hypermutation: low-affinity or autoreactive B-cell mutants are eliminated by apoptosis and the rest mature into memory B cells and long-lived plasma cells². Plasma cells might constitute an important component of B-cell memory, in particular those that recirculate to the bone marrow, where they are kept alive by as yet undefined stromal signals⁷³.

Although T cells can use CD95L to commit activation-induced suicide⁷, B cells generally do not express CD95L and die from a direct BCR-mediated signal. This opens the possibility that T cells kill CD95-positive B cells. This might apply to susceptible tolerant B cells

or to B cells insufficiently stimulated by survival signals or those whose BCRs are unoccupied by antigen⁷⁴.

Recently, the discovery of new receptor–ligand pairs within the TNF-R/TNF superfamilies has shed further light on the regulation of B-cell life and death^{9,72,75–77}. Blys (TALL-1, THANK, BAFF, zTNF4) and APRIL expressed on T cells and dendritic cells were found to bind to the receptors TACI and BCMA expressed on B cells upregulating nuclear factor (NF)- κ B, B-cell proliferation and immunoglobulin production. The receptor–ligand systems seem to act in concert to regulate B-cell function. Overstimulation of these systems can lead to autoimmunity and autoantibody formation, as in systemic lupus erythematosus. Blocking these systems might be used as a new treatment method in such diseases.

Interactions between APCs, T cells and B cells

T and B cells influence each other and influence persistence, clonal expansion and apoptosis of other cells. But it is the APCs that prime the T cells and initiate T-cell-dependent immunity⁴. APCs are able to engulf apoptotic and necrotic cells, and present their antigens to T cells (see review in this issue by Savill and Fadok, pages 784–788). But at present it is undecided whether material from apoptotic or necrotic cells activates or soothes the T cells^{78,79}. APCs are not passive bystander cells. Activated APCs synthesize CD95L, TRAIL, TNF and other factors that modulate the activity and function of T cells⁸⁰. In turn, activated T cells influence APC function and thereby affect the course of the immune response. At the initiation of the immune response, APCs must be resistant to apoptosis to exert their function⁸¹. Thus, switching these cells off to downregulate the response becomes an important issue. Two members of the TNF-R superfamily, CD40 and CD95, have adversarial roles in this context: the CD40–CD40L system allows the survival of APCs and the CD95–CD95L system induces their death⁸². The plasticity of the immune system might require that the cells can give and receive life and death signals at the same time and that it is the cellular context that determines which signal dictates the cellular response.

Diseases of the immune system involving apoptosis

Apoptosis is a fundamental process of regulation of the immune system; its derailment leads to severe diseases. Several examples of such diseases with either too little or too much apoptosis are discussed below.

Genetic defects in the CD95–CD95L system

Several mouse mutations have been identified that cause complex disorders of the immune system, manifested as lymphadenopathy and autoimmunity. One is the recessive *lpr* (lymphoproliferation) mutation. The symptoms of the disease arising from *lpr* are similar to those in systemic lupus erythematosus. The mutations *lpr*⁸³ (allelic to *lpr*) and *gld* (generalized lymphoproliferative disease) cause a very

similar disease. In all three cases, aberrant T cells accumulate; in *lpr* mice a splicing defect results in the greatly decreased expression of CD95. In *lpr*⁸³ mice a point mutation in the intracellular death domain of CD95 abolishes the transmission of the apoptotic signal. In *gld* mice a point mutation in the carboxy terminus of CD95L impairs its ability to interact successfully with its receptor. Thus, a failure of apoptosis accounts for the complex immune disorder in *lpr* and *gld* mutant mice⁸³.

In humans a similar disease with a dysfunction of the CD95 (type Ia ‘autoimmune lymphoproliferative syndrome’ (ALPS))–CD95L (type Ib ALPS) system has been reported. Children with ALPS (or Canale Smith syndrome) show massive, non-malignant lymphadenopathy, an altered and enlarged T-cell population and a severe autoimmunity. Many of these children show a crippling mutation in the death domain of CD95 but are heterozygous for this defect. Because the parents are unlike the children in being physically unaffected, a secondary, as yet unknown, defect must exist that is responsible for the symptoms^{24,25,84,85}.

In some cases (type II ALPS), defective CD95-mediated apoptosis is observed without mutations in CD95 or CD95L (ref. 86). This suggests that other defects that affect CD95 signalling exist; it has recently been suggested that mutations in caspase-10 might cause such a complementing defect.

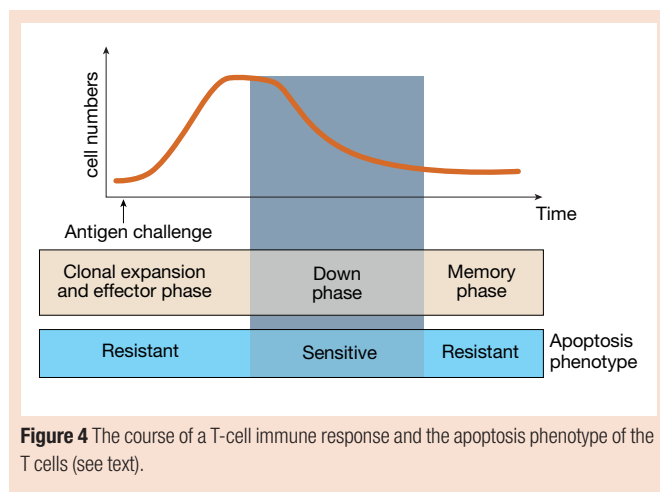
Thus, the inability of the immune system to eliminate self-reactive lymphocytes by apoptosis can cause autoimmunity. It is possible that autoantigen-driven prolonged T-cell activation might lock the cells into an AICD-resistant phenotype. Alternatively, apoptosis might produce changes in cellular constituents that affect antigen processing and self-tolerance. Increased resistance to apoptosis and persistence of autoreactive activated T cells have in fact been found in models of experimental autoimmunity. In these systems a clinically beneficial effect of drugs was observed that resensitized such T cells towards apoptotic deletion⁸⁷.

Apoptosis and lymphoid tumours

Apoptosis, or rather the lack of it, is important in the generation of tumours. Follicular lymphomas result from a translocation of *bcl-2* into the immunoglobulin heavy-chain locus and the deregulated expression of *bcl-2* under the influence of the immunoglobulin enhancer. Overexpression of *bcl-2* suppresses apoptosis and favours tumour cell proliferation⁸⁸. This is supported by the increased tumour incidence in, for example, *bcl-2* transgenic animals⁸⁹. Similarly, c-FLIP and most other apoptosis inhibitors might have oncogenic potential. In contrast, proapoptotic molecules could, in principle, act as tumour suppressors. Furthermore, resistance to chemotherapy might result from similar anti-apoptotic mechanisms to those known to block apoptosis in normal cells, and future therapeutic strategies will be aimed at sensitizing tumour cells to apoptosis while sparing normal tissue from drug damage⁹⁰ (see review in this issue by Nicholson, pages 810–816).

Tumours develop multiple mechanisms to evade elimination by the immune system. These mechanisms comprise a lack of expression of co-stimulatory or MHC molecules and active strategies such as the production of immunosuppressive cytokines. In addition, CD95L might have an immunosuppressive function. A number of tumours, including lymphoid tumours, are resistant to apoptosis and express functional CD95L constitutively or after chemotherapy. This situation might enable tumour cells to delete anti-tumour lymphocytes and to suppress anti-tumour immune responses, a phenomenon called ‘tumour counter-attack’.

CD95L is expressed constitutively in immune-privileged sites such as the testis and the eye and might contribute to the immune-privileged status by inducing apoptosis in infiltrating lymphocytes; this might be exploited in trying to delay the rejection of allografts. But it has also been reported that the overexpression of CD95L in grafts does not simply confer immune privilege; instead it induces a granulocytic response that accelerates rejection. Thus, the consequences of CD95L expression *in vivo* are far from clear, and further



experiments are needed to establish whether the counter-attack by tumours underlies their escape from the immune system *in vivo*. Once established, counter-attack could possibly be exploited for therapeutic use by either inhibiting it for cancer therapy or setting it up in organ transplantation⁹¹.

Apoptosis in AIDS

AIDS, characterized by a depletion of CD4⁺ T helper cells, is a disease with too much apoptosis. For example, the number of CD4⁺ T cells in the peripheral blood of individuals productively infected with HIV is low (in the range of one in several thousand). This implies that T-cell depletion in this disease also affects non-infected CD4⁺ T cells. How do they die? As discussed before, T cells can choose between several death-signalling pathways. These different signalling pathways might all be affected in AIDS. Further experiments are urgently needed to determine the contribution of such pathways to CD4⁺ T-cell depletion in this disease. The experiments described below discuss initial attempts to answer this question.

The regulatory viral gene products (for example, HIV-1 Tat) produced by HIV-infected cells penetrates non-infected cells and renders these cells hypersensitive to TCR-induced CD95-mediated apoptosis. HIV Tat induces a pro-oxidative state in the affected cells, increases CD95L expression and facilitates TCR-triggered CD95-mediated suicide. Further sensitization of the CD4⁺ T cells results from the binding of HIV gp120 to CD4 and from the cross-linking of bound gp120 by anti-gp120 antibodies^{92–98}. In addition to CD95-mediated apoptosis, a novel and rapid type of apoptosis induced by both HIV-binding cell-surface receptors, CD4 and CXCR4 (a chemokine receptor), was found in T-cell lines, human peripheral blood lymphocytes and CD4–CXCR4 transfectants. The potency of this phenomenon and its specificity for CD4⁺ T cells suggest that it might have a significant role in T helper cell depletion in AIDS. On the basis of these data, the use of antichemokine-receptor antibodies meant to prevent HIV-1 infection might be dangerous. But the use of the natural ligand of CXCR4, SDF-1 α , or its derivatives could be considered for therapy, because it inhibits infection as well as CXCR4-mediated apoptosis. Studying the apoptotic signalling cascade triggered by CD4 and CXCR4 might prove to be useful in the identification of therapeutic strategies aimed at intervening with the progressive loss of CD4⁺ T cells in HIV-1-infected individuals⁶¹.

Conclusion

Without death by apoptosis, the life of cells of the immune system and their precise and specific function would not be possible. We have begun to understand the signalling pathways of apoptosis in lymphocytes and the rules that determine lymphocyte interactions. We have also begun to understand the survival signals that counteract apoptosis and we might be capable, eventually, of manipulating the cells therapeutically (see review in this issue by Nicholson, pages 810–816). Finally, apoptosis adds new insight into the pathogenesis of diseases; this might be of therapeutic benefit in the future. □

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
2. Craxton, A., Otipoby, K. L., Jiang, A. & Clark, E. A. Signal transduction pathways that regulate the fate of B lymphocytes. *Adv. Immunol.* **73**, 79–152 (1999).
3. Sebзда, E. *et al.* Selection of the T cell repertoire. *Annu. Rev. Immunol.* **17**, 829–874 (1999).
4. Drakesmith, H., Chain, B. & Beverley, P. How can dendritic cells cause autoimmune disease? *Immunol. Today* **21**, 214–217 (2000).
5. DosReis, G. A. & Shevach, E. M. Peripheral T-cell self-reactivity and immunological memory. *Immunol. Today* **19**, 587–588 (1998).
6. Berzins, S. P., Godfrey, D. I., Miller, J. F. & Boyd, R. L. A central role for thymic emigrants in peripheral T cell homeostasis. *Proc. Natl Acad. Sci. USA* **96**, 9787–9791 (1999).
7. Krammer, P. H. CD95(APO-1/Fas)-mediated apoptosis: live and let die. *Adv. Immunol.* **71**, 163–210 (1999).
8. Peter, M. E., Scaffidi, C., Medema, J. P., Kischkel, F. C. & Krammer, P. H. In *Apoptosis, Problems and Diseases* (ed. Kumar, S.) 25–63 (Springer, Heidelberg, 1998).
9. Marsters, S. A. *et al.* Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* **10**, 785–788 (2000).
10. Cascino, I., Fiucci, G., Papoff, G. & Ruberti, G. Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing. *J. Immunol.* **154**, 2706–2713 (1995).
11. Klas, C., Debatin, K. M., Jonker, R. R. & Krammer, P. H. Activation interferes with the APO-1 pathway in mature human T cells. *Int. Immunol.* **5**, 625–630 (1993).
12. Leithäuser, F. *et al.* Constitutive and induced expression of Apo-1, a new member of the nerve growth factor/tumor necrosis factor receptor superfamily, in normal and neoplastic cells. *Lab. Invest.* **69**, 415 (1993).
13. Golstein, P. Controlling cell death. *Science* **275**, 1081–1082 (1997).
14. Medema, J. P. *et al.* Cleavage of FLICE (caspase-8) by granzyme B during cytotoxic T lymphocyte-induced apoptosis. *Eur. J. Immunol.* **27**, 3492–3498 (1997).
15. Martinez-Lorenzo, M. J. *et al.* Activated human T cells release bioactive Fas ligand and APO2 ligand in microvesicles. *J. Immunol.* **163**, 1274–1281 (1999).
16. Li, J. H. *et al.* The regulation of CD95 ligand expression and function in CTL. *J. Immunol.* **161**, 3943–3949 (1998).
17. Mariani, S. M., Matiba, B., Bäuml, C. & Krammer, P. H. Regulation of cell surface APO-1/Fas (CD95) ligand expression by metalloproteases. *Eur. J. Immunol.* **25**, 2303–2307 (1995).
18. Tanaka, M. *et al.* Fas ligand in human serum. *Nature Med.* **2**, 317–322 (1996).
19. Yagita, H. *et al.* Fas-mediated cytotoxicity — a new immunoregulatory and pathogenic function of Th1 CD4⁺ T cells. *Immunol. Rev.* **146**, 223–239 (1995).
20. Tanaka, M., Suda, T., Takahashi, T. & Nagata, S. Expression of the functional soluble form of human fas ligand in activated lymphocytes. *EMBO J.* **14**, 1129–1135 (1995).
21. Suda, T., Hashimoto, H., Tanaka, M., Ochi, T. & Nagata, S. Membrane Fas ligand kills human peripheral blood T lymphocytes, and soluble Fas ligand blocks the killing. *J. Exp. Med.* **186**, 2045–2050 (1997).
22. Kischkel, F. C. *et al.* Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.* **14**, 5579–5588 (1995).
23. Muzio, M. *et al.* FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
24. Siegel, R. M. *et al.* Fas preassociation required for apoptosis signaling and dominant inhibition by pathogenic mutations. *Science* **288**, 2354–2357 (2000).
25. Chan, F. K. *et al.* A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science* **288**, 2351–2354 (2000).
26. Scaffidi, C. *et al.* Two CD95 (APO-1/Fas) signaling pathways. *EMBO J.* **17**, 1675–1687 (1998).
27. Luo, X., Budihardjo, I., Zou, H., Slaughter, C. & Wang, X. Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* **94**, 481–490 (1998).
28. Li, H., Zhu, H., Xu, C. & Yuan, J. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* **94**, 491–501 (1998).
29. Kroemer, G. The pharmacology of T cell apoptosis. *Adv. Immunol.* **58**, 211–296 (1995).
30. Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
31. Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
32. Yin, X. M. *et al.* Bid-deficient mice are resistant to Fas-induced hepatocellular apoptosis. *Nature* **400**, 886–891 (1999).
33. Thome, M. *et al.* Viral Flc-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. *Nature* **386**, 517–521 (1997).
34. Hu, S. *et al.* A novel family of viral death effector domain-containing molecules that inhibit both CD-95- and tumor necrosis factor receptor-1-induced apoptosis. *J. Biol. Chem.* **272**, 9621–9624 (1997).
35. Bertin, J. *et al.* Death effector domain-containing herpesvirus and poxvirus proteins inhibit both Fas- and TNFR1-induced apoptosis. *Proc. Natl Acad. Sci. USA* **94**, 1172–1176 (1997).
36. Yeh, W. C. *et al.* Requirement for Casper (c-FLIP) in regulation of death receptor-induced apoptosis and embryonic development. *Immunity* **12**, 633–642 (2000).
37. Bodmer, J. L. *et al.* TRAIL receptor-2 signals apoptosis through FADD and caspase-8. *Nature Cell Biol.* **2**, 241–243 (2000).
38. Sprick, M. R. *et al.* FADD/MORT1 and caspase-8 are recruited to TRAIL receptors 1 and 2 and are essential for apoptosis mediated by TRAIL receptor 2. *Immunity* **12**, 599–609 (2000).
39. Kischkel, F. C. *et al.* Apo2L/TRAIL-dependent recruitment of endogenous FADD and caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611–620 (2000).
40. Chen, W. F., Scollay, R., Clark-Lewis, I. & Shortman, K. The size of functional T-lymphocyte pools within thymic medullary and cortical cell subsets. *Thymus* **5**, 179–195 (1983).
41. Surh, C. D. & Sprent, J. T-cell apoptosis detected *in situ* during positive and negative selection in the thymus. *Nature* **372**, 100–103 (1994).
42. Amsen, D. & Kruisbeek, A. M. Thymocyte selection: not by TCR alone. *Immunol. Rev.* **165**, 209–229 (1998).
43. Kishimoto, H. & Sprent, J. The thymus and central tolerance. *Clin. Immunol.* **95**, S3–S7 (2000).
44. Newton, K., Harris, A. W. & Strasser, A. FADD/MORT1 regulates the pre-TCR checkpoint and can function as a tumour suppressor. *EMBO J.* **19**, 931–941 (2000).
45. Krammer, P. H. & Debatin, K. M. When apoptosis fails. *Curr. Biol.* **2**, 383–385 (1992).
46. Kishimoto, H., Surh, C. D. & Sprent, J. A role for Fas in negative selection of thymocytes *in vivo*. *J. Exp. Med.* **187**, 1427–1438 (1998).
47. Aiello, S. *et al.* Thymic dendritic cells express inducible nitric oxide synthase and generate nitric oxide in response to self- and alloantigens. *J. Immunol.* **164**, 4649–4658 (2000).
48. Nelson, B. H. & Willerford, D. M. Biology of the interleukin-2 receptor. *Adv. Immunol.* **70**, 1–81 (1998).
49. Dhein, J., Walczak, H., Bäuml, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by APO-1/(Fas/CD95). *Nature* **373**, 438–441 (1995).
50. Alderson, M. R. *et al.* Fas ligand mediates activation-induced cell death in human T lymphocytes. *J. Exp. Med.* **181**, 71–77 (1995).
51. Brunner, T. *et al.* Cell-autonomous Fas (CD95)/Fas-ligand interaction mediates activation-induced apoptosis in T-cell hybridomas. *Nature* **373**, 441–444 (1995).
52. Ju, S. T. *et al.* Fas(CD95)/FasL interactions required for programmed cell death after T-cell activation. *Nature* **373**, 444–448 (1995).
53. Van Parijs, L., Ibragimov, A. & Abbas, A. K. The roles of costimulation and Fas in T cell apoptosis and peripheral tolerance. *Immunity* **4**, 321–328 (1996).
54. Singer, G. G. & Abbas, A. K. The fas antigen is involved in peripheral but not thymic deletion of T lymphocytes in T cell receptor transgenic mice. *Immunity* **1**, 365–371 (1994).
55. Inaba, M. *et al.* Primed T cells are more resistant to Fas-mediated activation-induced cell death than naive T cells. *J. Immunol.* **163**, 1315–1320 (1999).

56. Van Parijs, L. *et al.* Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. *Immunity* **11**, 281–288 (1999).
57. Zheng, L. *et al.* Induction of apoptosis in mature T cells by tumour necrosis factor. *Nature* **377**, 348–351 (1995).
58. Zimmerman, C., Brduscha-Riem, K., Blaser, C., Zinkernagel, R. M. & Pircher, H. Visualization, characterization, and turnover of CD8⁺ memory T cells in virus-infected hosts. *J. Exp. Med.* **183**, 1367–1375 (1996).
59. Hildeman, D. A. *et al.* Reactive oxygen species regulate activation-induced T cell apoptosis. *Immunity* **10**, 735–744 (1999).
60. Spaner, D., Raju, K., Rabinovich, B. & Miller, R. G. A role for perforin in activation-induced T cell death *in vivo*: increased expansion of allogeneic perforin-deficient T cells in SCID mice. *J. Immunol.* **162**, 1192–1199 (1999).
61. Berndt, C., Möppts, B., Angermüller, S., Gierschik, P. & Krammer, P. H. CXCR4 and CD4 mediate a rapid CD95-independent cell death in CD4⁺ T cells. *Proc. Natl Acad. Sci. USA* **95**, 12556–12561 (1998).
62. Kirchhoff, S., Müller, W. W., Li-Weber, M. & Krammer, P. H. Upregulation of c-FLIP and reduction of AICD in CD28 costimulated human T cells. *Eur. J. Immunol.* (in the press).
63. Kirchhoff, S., Müller, W. W., Krueger, A., Schmitz, I. & Krammer, P. H. TCR-mediated upregulation of c-FLIP, correlates with resistance towards CD95 mediated apoptosis by blocking DISC activity. *J. Immunol.* (in the press).
64. Van Parijs, L., Refaeli, Y., Abbas, A. K. & Baltimore, D. Autoimmunity as a consequence of retrovirus-mediated expression of C-FLIP in lymphocytes. *Immunity* **11**, 763–770 (1999).
65. Boise, L. H., Noel, P. J. & Thompson, C. B. CD28 and apoptosis. *Curr. Opin. Immunol.* **7**, 620–625 (1995).
66. Lam, K. P. & Rajewsky, K. Rapid elimination of mature autoreactive B cells demonstrated by Cre-induced change in B cell antigen receptor specificity *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 13171–13175 (1998).
67. Lam, K. P., Kuhn, R. & Rajewsky, K. *In vivo* ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* **90**, 1073–1083 (1997).
68. Bouchon, A., Krammer, P. H. & Walczak, H. Critical role for mitochondria in B cell receptor-mediated apoptosis. *Eur. J. Immunol.* **30**, 69–77 (2000).
69. Berard, M. *et al.* Mitochondria connects the antigen receptor to effector caspases during B cell receptor-induced apoptosis in normal human B cells. *J. Immunol.* **163**, 4655–4662 (1999).
70. Lagresle, C., Mondiere, P., Bella, C., Krammer, P. H. & Defrance, T. Concurrent engagement of CD40 and the antigen receptor protects naive and memory human B cells from APO-1/Fas-mediated apoptosis. *J. Exp. Med.* **183**, 1377–1388 (1996).
71. Smith, K. G. *et al.* *bcl-2* transgene expression inhibits apoptosis in the germinal center and reveals differences in the selection of memory B cells and bone marrow antibody-forming cells. *J. Exp. Med.* **191**, 475–484 (2000).
72. Knodel, M., Kuss, A. W., Lindemann, D., Berberich, I. & Schimpl, A. Reversal of Blimp-1-mediated apoptosis by A1, a member of the Bcl-2 family. *Eur. J. Immunol.* **29**, 2988–2998 (1999).
73. Merville, P. *et al.* Bcl-2⁺ tonsillar plasma cells are rescued from apoptosis by bone marrow fibroblasts. *J. Exp. Med.* **183**, 227–236 (1996).
74. Foote, L. C., Marshak-Rothstein, A. & Rothstein, T. L. Tolerant B lymphocytes acquire resistance to Fas-mediated apoptosis after treatment with interleukin 4 but not after treatment with specific antigen unless a surface immunoglobulin threshold is exceeded. *J. Exp. Med.* **187**, 847–853 (1998).
75. Xia, X. Z. *et al.* TACI is a TRAF-interacting receptor for TALL-1, a tumor necrosis factor family member involved in B cell regulation. *J. Exp. Med.* **192**, 137–144 (2000).
76. Gross, J. A. *et al.* TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease. *Nature* **404**, 995–999 (2000).
77. Shu, H. B. & Johnson, H. B cell maturation protein is a receptor for the tumor necrosis factor family member TALL-1. *Proc. Natl Acad. Sci. USA* **97**, 9156–9161 (2000).
78. Steinman, R. M., Turley, S., Mellman, I. & Inaba, K. The induction of tolerance by dendritic cells that have captured apoptotic cells. *J. Exp. Med.* **191**, 411–416 (2000).
79. Albert, M. L., Sauter, B. & Bhardwaj, N. Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs. *Nature* **392**, 86–89 (1998).
80. Fanger, N. A., Maliszewski, C. R., Schooley, K. & Griffith, T. S. Human dendritic cells mediate cellular apoptosis via tumor necrosis factor-related apoptosis-inducing ligand (TRAIL). *J. Exp. Med.* **190**, 1155–1164 (1999).
81. Ashany, D., Savir, A., Bhardwaj, N. & Elkon, K. B. Dendritic cells are resistant to apoptosis through the Fas (CD95/APO-1) pathway. *J. Immunol.* **163**, 5303–5311 (1999).
82. Björck, P., Banchereau, J. & Flores-Romo, L. CD40 ligation counteracts Fas-induced apoptosis of human dendritic cells. *Int. Immunol.* **9**, 365–372 (1997).
83. Nagata, S. Human autoimmune lymphoproliferative syndrome, a defect in the apoptosis-inducing Fas receptor: a lesson from the mouse model. *J. Hum. Genet.* **43**, 2–8 (1998).
84. Fisher, G. H. *et al.* Dominant interfering Fas gene mutations impair apoptosis in a human autoimmune lymphoproliferative syndrome. *Cell* **81**, 935–946 (1995).
85. Rieux-Laucat, F. *et al.* Mutations in Fas associated with human lymphoproliferative syndrome and autoimmunity. *Science* **268**, 1347–1349 (1995).
86. Wang, J. *et al.* Inherited human Caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type II. *Cell* **98**, 47–58 (1999).
87. Zhou, T. *et al.* Bisindolylmaleimide VIII facilitates Fas-mediated apoptosis and inhibits T cell-mediated autoimmune diseases. *Nature Med.* **5**, 42–48 (1999).
88. Chao, D. T. & Korsmeyer, S. J. Bcl-2 family: regulators of cell death. *Annu. Rev. Immunol.* **16**, 395–419 (1998).
89. Adams, J. M., Harris, A. W., Strasser, A., Ogilvy, S. & Cory, S. Transgenic models of lymphoid neoplasia and development of a pan-hematopoietic vector. *Oncogene* **18**, 5268–5277 (1999).
90. Debatin, K. M. Activation of apoptosis pathways by anticancer drugs. *Adv. Exp. Med. Biol.* **457**, 237–244 (1999).
91. Igney, F. H., Behrens, C. K. & Krammer, P. H. Tumor counterattack — concept and reality. *Eur. J. Immunol.* **30**, 725–731 (2000).
92. Westendorp, M. O. *et al.* Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. *Nature* **375**, 497–500 (1995).
93. Krammer, P. H. *et al.* The role of APO-1-mediated apoptosis in the immune system. *Immunol. Rev.* **142**, 175–191 (1994).
94. Debatin, K. M. *et al.* High expression of APO-1 (CD95) on T lymphocytes from human immunodeficiency virus-1-infected children. *Blood* **83**, 3101–3103 (1994).
95. Baumler, C. B. *et al.* Activation of the CD95 (APO-1/Fas) system in T cells from human immunodeficiency virus type-1-infected children. *Blood* **88**, 1741–1746 (1996).
96. Li, C. J., Friedman, D. J., Wang, C., Metelev, V. & Pardee, A. B. Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein. *Science* **268**, 429–431 (1995).
97. Finkel, T. H. *et al.* Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. *Nature Med.* **1**, 129–134 (1995).
98. Gougeon, M. L. & Montagnier, L. Programmed cell death as a mechanism of CD4 and CD8 T cell deletion in AIDS. Molecular control and effect of highly active anti-retroviral therapy. *Ann. N.Y. Acad. Sci.* **887**, 199–212 (1999).

Acknowledgements

I thank my previous and present collaborators for help with the manuscript and the figures, particularly S. Baumann, K.-M. Debatin, T. Defrance, C. S. Falk, S. Kirchhoff, A. Krueger, B. Kyewski, M. Peter, I. Schmitz, A. Strecker and H. Walczak. I also thank H. Sauter and B. Pétillon for secretarial assistance. This work was funded by Deutsche Krebshilfe Dr Mildred Scheel Stiftung, German Israeli Cooperation in Cancer Research, AIDS grant German Federal Health Agency, Tumor Centre Heidelberg/Mannheim, BMBF Förderschwerpunkte 'Clinical-biomedical research' and 'Apoptosis', AIDS Verbund Heidelberg, Wilhelm-Sander Stiftung, Ernst-Jung-Stiftung, Förderschwerpunkt Transplantation, and the DFG. I apologize to all my colleagues who have done excellent work in the field and whose papers have not been quoted comprehensively — it was not possible to be encyclopedic in this exponentially growing field. I dedicate this article to the Basel Institute for Immunology, the closure of which was recently announced.

Apoptosis in development

Pascal Meier^{*‡}, Andrew Finch[†] & Gerard Evan[†]

^{*}Signal Transduction Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

[†]UCSF Cancer Center, 2340 Sutter Street, San Francisco, California 94143-0875, USA (e-mail: gevan@cc.ucsf.edu; afinch@cc.ucsf.edu)

[‡]Present address: The Breakthrough Toby Robins Breast Cancer Research Centre, Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK (e-mail: meierp@icr.ac.uk)

Essential to the construction, maintenance and repair of tissues is the ability to induce suicide of supernumerary, misplaced or damaged cells with high specificity and efficiency. Study of three principal organisms — the nematode, fruitfly and mouse — indicate that cell suicide is implemented through the activation of an evolutionarily conserved molecular programme intrinsic to all metazoan cells. Dysfunctions in the regulation or execution of cell suicide are implicated in a wide range of developmental abnormalities and diseases.

Those interested in the 'Big Dig', the city of Boston's heroic attempt to bury Interstate 93 beneath its pavements while maintaining a passable stab at business-as-usual above, will be well acquainted with the idea that major construction entails a substantial amount of demolition. So too in animal development: during the ontogeny of many organs, cells are over-produced only to be etched or whittled away to generate the rococo structures of functional tissues. Early distaste among biologists for the 'wastefulness' of such a process has given away to the recognition that the ability to ablate cells is as essential a constructive process in animal ontogeny as are the abilities to replicate and differentiate them. After all, most animals thrive in a sea of energy and profligacy with their component cells is a small price to pay for ability to move around and propagate. It is highly unlikely that the peacock, upon encountering the peahen of his dreams, demurs to ponder the energetic cost of his outrageous tail.

It is now clear that physiological cell death is an essential component of animal development, important for establishment and, in vertebrates at least, maintenance of tissue architecture. A general *modus operandi* of metazoan development is the over-production of excess cells followed by an apoptotic culling during later stages of development to match the relative number of cells of different types to achieve proper organ function¹. Thus, during animal development, numerous structures are formed that are later removed by apoptosis. This enables greater flexibility as primordial structures can be adapted for different functions at various stages of life or in different sexes. Thus, the Müllerian duct gives rise to the uterus and oviduct in females but it is not needed in males and so is consequently removed. On the other hand, the Wolffian duct is the source of male reproductive organs and is deleted in females. Organisms are like many modern computer programs, full of remnant code that was once used in an ancestral incarnation or that runs irrelevant routines that nobody needs. During development, apoptosis is frequently used to expunge such structures. For instance, early in vertebrate development, the pronephric kidney tubules arise from the nephrogenic mesenchyme. Although these pronephric tubules form functioning kidneys in fish and in amphibian larvae, they are not active in mammals and degenerate². Similarly, during insect and amphibian metamorphosis, apoptosis ablates cells that are no longer needed such as muscles and neurons essential for larval locomotion in insects or the amphibian tadpole tail.

Apoptosis also acts as part of a quality-control and repair mechanism that contributes to the high level of plasticity during development by compensating for many genetic or stochastic developmental errors. For example, *Drosophila* embryos with extra doses of the morphogen bicoid (*bcd*) gene show severe mispatterning in their anterior regions. Surprisingly, these embryos develop into relatively normal larvae and adults because cell death compensates for tissue overgrowth and mispatterning³. Cells that have been incorrectly programmed are, in effect, misplaced cells. They therefore fail to receive the appropriate trophic signals for their survival and consequently activate their innate auto-destruct mechanism.

In this review we outline how each of the three great model organisms of developmental biology, the nematode *Caenorhabditis elegans*, the fruitfly *Drosophila melanogaster* and the mouse *Mus musculus* have contributed to our understanding of the role of cell death in development and homeostasis.

Lessons from invertebrates

The first evidence for a genetic basis of apoptosis came from studies in *C. elegans* whose invariant, lineage-restricted development makes this organism particularly advantageous for the study of developmental processes. During ontogeny of the adult hermaphrodite worm, 131 of the 1,090 somatic cells die by apoptosis, leaving an adult comprised of 959 cells. Genetic screens for mutants defective in cell death identified specific genes required for regulation, execution and resolution of apoptosis, of which four, *egl-1*, *ced-3*, *ced-4* and *ced-9*, are required for each cellular demise (Fig. 1). Loss-of-function mutations in *egl-1*, *ced-3* or *ced-4* result in survival of all 131 doomed cells, implicating these three genes in the induction of cell death. In contrast, animals lacking functional *ced-9* die early in development owing to massive ectopic cell death, whereas a gain-of-function mutation in *ced-9* blocks all 131 cell deaths, implicating *ced-9* as a suppressor of cell death.

Remarkably, this basic cell death machinery is highly conserved throughout metazoan evolution. *ced-3* encodes CED-3, a cysteine protease of an evolutionarily conserved class now dubbed 'caspases' because of their predilection for cleaving at aspartyl residues. By their cleavage of critical cellular substrates, caspases act as key engines of cellular destruction in all metazoans⁴ (see review in this issue by Hengartner, pages 770–776). Like most proteases, caspases are synthesized as pro-enzyme zymogens that have little or no intrinsic catalytic activity. They are activated by

proteolytic cleavage either through the action of other caspases or through an autocatalytic process in which multiple procaspase molecules are brought into close proximity through formation of multiprotein 'apoptosome' complexes⁵. Such complexes permit the low intrinsic proteolytic activity of the procaspases to trigger their own intermolecular cleavage and activation. In addition to CED-3, two other *C. elegans* caspases have been identified, CSP-1 and CSP-2 (ref. 6). However, the lack of cell death in *ced-3* mutants indicates that neither can replace CED-3 function and it is probable that both act as part of a downstream proteolytic cascade.

Caspases can be grouped into two general types based on the size of their amino-terminal prodomains. Caspases with short prodomains (type 2) are in general activated by upstream caspase cleavage and act as 'effectors' that implement apoptosis by cleavage of appropriate substrates. In contrast, the extended prodomains of the so-called type 1 'initiator' caspases, of which CED-3 is an example, serve as interaction domains for assembly into 'apoptosome' complexes, an assembly that is dependent on specific adaptor or scaffolding molecules and typically occurs in response to activation of some pro-apoptotic signalling pathways. In *C. elegans*, the requisite adaptor protein is encoded by *ced-4*, although its innate ability to trigger CED-3 activation is stanchied by the protein product of the *ced-9* death suppressor gene. Only when CED-4 is displaced from CED-9 by the EGL-1 protein is the lethal proteolytic action of CED-3 unleashed to ablate its 131 cellular victims. Evidence indicates that EGL-1 can be regulated transcriptionally. For example, EGL-1 expression induces apoptosis in hermaphrodite-specific neurons of male worms whereas its expression in hermaphrodites is repressed by the TRA-1A sex determination transcription factor⁷. Although EGL-1 and the CED proteins are implicated in all developmental cell deaths in *C. elegans*, not all cell deaths are regulated in the same way. For example, CES-1 and CES-2 act to regulate apoptosis in specific neurons. CES-1 is an anti-apoptotic zinc-finger transcriptional repressor of the Snail/Slug family⁸ whose apoptotic target genes are unknown⁹. CES-2 is a PAS bZip protein, related to mammalian hepatocyte leukaemia factor¹⁰, that acts to promote apoptosis through repression of CES-1 expression¹¹. Another cell type-specific example is the death of germ cells in the hermaphrodite gonad, which uses CED-3, CED-4 and CED-9 but is independent of EGL-1. This example of nematode apoptosis is also interesting because it is not pre-programmed but occurs in an adaptive way in response to DNA damage, age and environmental factors and is modulated by the Ras/mitogen-activated protein kinase (MAPK) pathway (see below)^{12,13}.

The general mechanistic interplay of the *C. elegans* cell death machinery is conserved, albeit with substantial embellishments, in other metazoans. Multiple caspases are present in both *Drosophila* and mammals, and these are in turn regulated by various homologues and analogues of the CED-4 adaptor/scaffold protein of which the evolutionarily closest known functional relatives are Apaf-1 in man¹⁴ and dApaf-1/DARK/HAC-1 in flies¹⁵⁻¹⁷. In addition, in mammals at least, certain caspases are activated by recruitment into complexes induced by ligation of death receptors such as CD95 (Apo-1/Fas) and tumour necrosis factor (TNF) receptor 1 (see review in this issue by Krammer, pages 789-795). The mammalian¹⁸ and recently identified *Drosophila*¹⁹⁻²² counterparts of the CED-9 death suppressor protein are the Bcl-2 family of proteins which, in mammals (and maybe in *Drosophila*), includes both anti-apoptotic ('CED-9/Bcl-2-like') and BH3 pro-apoptotic ('EGL-1-like') members²³. The remarkable conservation of molecular mechanism by which Bcl-2 proteins prevent cell death is most graphically demonstrated by the fact that human Bcl-2 is fully functional in suppressing cell deaths in developing *C. elegans*²⁴.

Developmental cell death in *Drosophila*

It is a matter of debate whether *C. elegans* exemplifies an evolutionarily simple prototypic organism or a highly compressed and stripped down version of a more complex one. Whichever, its apoptotic machinery is clearly adapted to implementing cell death in a

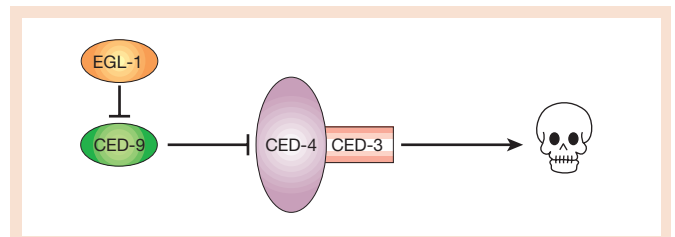


Figure 1 The apoptotic system in *C. elegans*. When compared with the apoptotic systems in *D. melanogaster* (Fig. 2) and mammals (Fig. 3) it is clear that fundamental components of the apoptotic pathways are conserved, but an increasing complexity from *C. elegans* to mammals is apparent. In Figs 1-3, proteins that are homologous (by sequence or functionally) between the three organisms are similarly depicted. Prototypic molecules are used to represent families of proteins.

highly invariant manner. In contrast, development of *D. melanogaster* evidences a far greater complexity and plasticity which is mirrored in an apoptotic machinery that, through evolutionary duplication and elaboration, is significantly more complex than that of its nematode cousin (Fig. 2). Intriguingly, this greater complexity includes a cadre of cell death regulatory proteins that so far has not been found in either nematode or mammal. Genetic analysis of *Drosophila* developmental cell death indicates that it is in the main determined by three pro-apoptotic proteins: reaper (RPR), GRIM and head involution defective (HID). Embryos bearing the *Df(3)H99* deletion in chromosome 3 that ablates the *rpr*, *grim* and *hid* loci show essentially no embryonic apoptosis and die towards the end of embryogenesis with a significant excess of cells²⁵. Expression of two of these 'death' genes, *rpr* and *grim*, is confined only to those cells destined to die: their expression presages death by some two hours. However, the *rpr* gene also triggers apoptosis in an adaptive way. *rpr* is induced in response to developmental malfunction, although the mechanism is unclear, and it is also a transcriptional target of the *Drosophila* p53 protein²⁶⁻²⁸, making its expression responsive to genotoxic stress²⁸. Inducible activation of *rpr* therefore provides *Drosophila* with an adaptive mechanism for specifically ablating misplaced and genetically damaged cells.

Drosophila apoptosis is also regulated critically by survival signals provided by neighbouring cells. Thus, HID, in contrast to RPR and GRIM, is expressed in cells that both live and die but its action is regulated by the Ras/Raf/MAPK signalling pathway; this pathway promotes survival both by downregulating HID expression and by inactivating the existing HID protein through phosphorylation^{29,30}. Nonetheless, because genetic analyses indicate that *rpr*, *grim* and *hid* act cooperatively to induce apoptosis in a cell-specific manner, modulation of HID is only one of several factors that determine cell viability in any tissue. Furthermore, at least some cell death in *Drosophila* development is independent of *rpr*, *grim* and *hid*: for example, nurse-cell apoptosis during oogenesis occurs normally in *H99* deletion flies³¹.

Whether or not RPR, GRIM and HID have any direct homologues in mammals, they nonetheless interact with components of *Drosophila* cell death machinery that are conserved, in particular the evolutionarily conserved inhibitor-of-apoptosis protein (IAP) family that in mammals are targets of the pro-apoptotic Smac/DIABLO protein³²⁻³⁴. Loss-of-function mutations in the *Drosophila* IAP DIAP-1 result in embryonic death as a result of extensive apoptosis, whereas ectopic expression of the IAPs DIAP1, DIAP2 or DETERIN suppresses cell death induced by RPR, GRIM or HID. Although some IAPs have been shown to act as direct competitive inhibitors of caspases, it seems likely that many act to bind to the large prodomains of type 1 caspases and thereby prevent their sequestration into activating apoptosome complexes. It is thought that RPR, HID and GRIM inhibit such interactions between IAPs and caspases, so promoting caspase activation and cell death³⁵⁻³⁸. In this way, IAPs act

as 'guardians' of the cell death machinery. However, in cells that are fated to die, this IAP-mediated road block has to be overcome and it seems that RPR, GRIM and HID function to do this, at least in part, by antagonizing the anti-apoptotic activity of DIAP1, thereby liberating caspases.

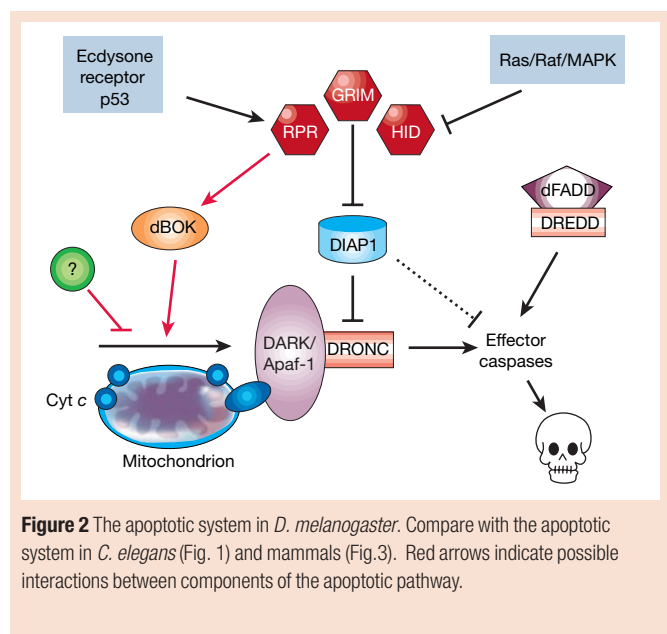
An ever growing number of Bcl-2 family members is emerging in both *Drosophila* and mammals, each member characterized by its ability to either induce or to suppress apoptosis¹⁸. Bcl-2 family proteins form homo- and heterodimers and it seems to be the net balance of protectors and killers that determines whether a cell is to live or die. It is probable that the recently identified *Drosophila* pro-apoptotic members of the Bcl-2 family such as dBORG-1/DROB-1/DEBCL/dBOK and dBORG-2/BUFFY proteins (reviewed in ref. 39) are important in determining cell survival during development. Loss of dBORG-1 function results in excess glial cells, attesting to the protein's pro-apoptotic activity^{19–22}, but in other circumstances dBORG-1 seems to exert a protective effect. Unfortunately, too little is currently known of either the biochemistry or genetics of the various *Drosophila* Bcl-2 homologues to deduce their roles during fly development.

Caspases are clearly necessary for *Drosophila* cell death as their inhibition by IAPs, the baculovirus p35 caspase inhibitor or dominant-negative caspase mutants inhibits developmental apoptosis as well as apoptosis induced by overexpression of *rpr*, *hid* or *grim* (reviewed in ref. 40). Five caspases have been identified in *Drosophila* — DCP-1, drICE, DCP-2/DREDD, DRONC and DECAY — with another two predicted on the basis of genomic sequence. Both DCP-2/DREDD and DRONC have extensive prodomains suggesting that they are initiator caspases linked to specific upstream apoptotic signalling pathways. In contrast, DCP-1, drICE and DECAY all possess short prodomains and probably act as downstream 'effector' caspases that are activated by upstream initiator caspases. The absence of loss-of-function mutants of most *Drosophila* caspases makes it difficult to determine their various biological role and level of redundancy. However, loss of zygotic DCP-1 leads to absence of gonads or imaginal discs (epithelial structures that will give rise to the adult insect): such animals have brittle tracheae, exhibit prominent melanotic tumours and die during larval stages. Moreover, female chimaeras bearing *dcp*^{-/-} germline cells are sterile owing to a failure of nurse cells to reorganize their actin cytoskeletons, an essential process for the cytoplasmic transfer from nurse cells into the oocyte⁴¹.

Huge numbers of cells die during *Drosophila* embryonic and imaginal development, as well as during metamorphosis. Throughout metamorphosis, pulses of the steroid hormone 20-hydroxyecdysone (ecdysone) trigger the transition from larval to adult life and signal stage- and tissue-specific onset of apoptosis. During this transition phase the larva is profoundly reorganized to build the adult insect. Most structures that are no longer needed in the adult are deleted by apoptosis while others are newly built from imaginal precursor cells. For example, destruction of the salivary glands during metamorphosis is triggered by an ecdysone-initiated switch in gene expression whereby *DIAP2* expression is downregulated and *rpr* and *hid* expression are both induced⁴². *rpr* induction occurs by means of an ecdysone-receptor response element in the *rpr* promoter. The *rpr* promoter/enhancer element, which also contains a p53-responsive element (see above), is extensive and contains a plethora of different response elements that serve to integrate *rpr* expression into a variety of pro-apoptotic signalling pathways.

Apoptosis in vertebrates

In both *C. elegans* and *Drosophila*, development is restricted largely to early life and ends at birth or metamorphosis. In vertebrates, by contrast, the developmental processes of morphogenesis, remodelling and regeneration are sustained at high level in many tissues, either constitutively or in response to insult or injury, throughout their extensive life spans. The critical role of apoptosis in development is therefore evident throughout vertebrate life and, consequently, dysfunctions in apoptosis manifest themselves not only in develop-



mental abnormalities but also in a wide variety of adult pathologies. It is these pathologies, in particular cancer (see review in this issue by Rich *et al.*, pages 777–783) and degenerative disease (see review by Yuan and Yankner, pages 802–809), that have yielded much information as to the variegated roles of apoptosis in vertebrate biology.

Vertebrate apoptotic machinery is substantially homologous to that of invertebrates, although it is more elaborate and degenerate: caspases, Bcl-2 and IAP family proteins, and survival signalling pathways all exist in bewildering multiplicity, consistent with the more sophisticated needs for control of apoptosis in vertebrate tissues (Fig. 3, and see below). As in invertebrates, a variety of transcriptional mechanisms is important in the regulation of developmental apoptosis in vertebrates. For example, regression of the tadpole tail during metamorphosis requires thyroid hormone-induced RNA and protein synthesis for its destruction⁴³, steroid hormone receptors are critical controllers of apoptosis in many mammalian tissues including the mammary gland, the prostate, the ovary and testis (reviewed in ref. 44), and the classical apoptotic paradigm of interdigital cell death is determined by the transcriptional readout of the transforming growth factor- β signalling pathway (reviewed in ref. 45). Nonetheless, mammalian apoptosis is also significantly regulated, in both development and throughout life, by two non-transcriptional signalling systems whose exact counterparts are either absent or have proven elusive in the fly and worm.

First, mammals possess a family of death receptors whose ligation can directly trigger activation of specific initiator caspases through induced assembly of discrete apoptosome complexes. The archetypal members of this death-receptor family are the TNF and CD95 receptors that recruit caspase-8 via the adaptor protein FADD (for Fas-associated death domain protein; see review in this issue by Krammer, pages 789–795). Among other things, mammalian death receptors are used by cytotoxic T lymphocytes to impose an incontestable cell death programme upon target infected cells. Although no equivalent of death receptors has yet been identified in *Drosophila*, a homologue of the FADD adaptor has recently been isolated that interacts with the prodomain of the apical caspase DCP-2/DREDD (ref. 46). Furthermore, a suspicion that *Drosophila* FADD may, as in mammals, be linked to some kind of death receptor is fostered by the observation that expression of mammalian CD95 in insect cells induces apoptosis⁴⁷.

Second, in mammals many pro-apoptotic insults seem to impact directly upon mitochondria to induce their leakiness and the release of various pro-apoptotic polypeptides. One of these is

holocytochrome *c*, which orchestrates assembly of a complex involving Apaf-1, the closest known mammalian homologue of nematode CED-4, with caspase-9, a CARD (caspase-activating recruitment domain) initiator caspase that is then activated and triggers a downstream cascade of effector caspases^{14,48}. Another is Smac/DIABLO, which binds and antagonizes the anti-apoptotic activity of XIAP (refs 32, 33) and is probably the functional analogue of *Drosophila* RPR, HID and GRIM (ref. 34). In mammals, the principal anti-apoptotic action of Bcl-2 proteins, and of the survival signalling pathways that impact on them, seems to be the stabilization of mitochondrial integrity and prevention of release of these pro-apoptotic polypeptides. This differs from the role of nematode CED-9, which acts by directly interfering with the ability of CED-4 to activate CED-3. Indeed, there is no evidence for any involvement of mitochondria in cell death in *C. elegans*. The *Drosophila* CED-4/Apaf-1 adaptor is far more similar to Apaf-1 of mammals than to CED-4 and there is clear evidence of a role for cytochrome *c* in fly apoptosis^{17,49}. Thus, so far it remains unclear whether insects and vertebrates have evolved a more complex elaboration of the apoptotic machinery that incorporates the mitochondrion or whether *C. elegans* represents a stripped down version of a more evolutionarily ancient mechanism.

For all their apparent sophistication, vertebrate tissues are constructed using the same three general principles as worms and flies – cell proliferation, differentiation and demolition. However, among other factors the size and longevity of vertebrates place peculiar demands on their apoptotic machinery. For example, although in general most vertebrate tissues may be considered no more intrinsically complex than those of invertebrates, merely larger, there are two notable exceptions. The intricacies of the vertebrate central nervous and immune systems are (with the possible exception of the sadly under-investigated cephalopods) without peer in the animal world. Both of these tissues self-assemble through an extensive iterative matching process which is governed by application of fairly simple genetically programmed rules rather than through implementation of any specific cellular map of the final organ. Such matching is an evolutionarily ancient method of construction used, for example, in the developing fly nervous system, but its extent in vertebrates is without precedent. It is also an imprecise and stochastic affair that generates vast numbers of unmatched orphan cells that must be efficiently culled and removed to enable productive networks to emerge. The solution to this problem lies in configuring component cells in each tissue to commit suicide unless they establish productive connections. Thus, more than 80% of ganglion cells in the cat retina die shortly after they are born because their survival depends on the availability of limiting amounts of neurotrophic factors secreted by the target cells they innervate and for which they compete. A similar selective attrition occurs during development of the optic nerve⁵⁰ and, to various extents, in the entire central and peripheral nervous systems. In all tested cases, cell death can be largely suppressed by the excess provision of an appropriate neurotrophic survival factor. In the vertebrate immune system, cell ‘wastage’ is even more profound: survival of the emerging lymphocyte is absolutely dependent upon the fickle assembly of a productive immune receptor that provides the trophic signal necessary to suppress apoptosis. In this way, lymphocytes bearing inoperative or self-reactive receptors are deleted from the immune repertoire.

Largely on the basis of such studies, Martin Raff first postulated that cell death is the default state of all metazoan cells which must be continuously forestalled by environmental survival signals^{51,52}. Examples of survival signals include soluble cytokines and hormones, synaptic connections, and direct physical interactions with heterotypic cell neighbours and extracellular matrix. Different cell types require differing combinations of survival signals, which are only available within discrete somatic environments. Consequently, somatic cells are to great degree ‘trapped’ within specialized microenvironments within the body, dying should they stray or become dislodged through injury, chance or developmental

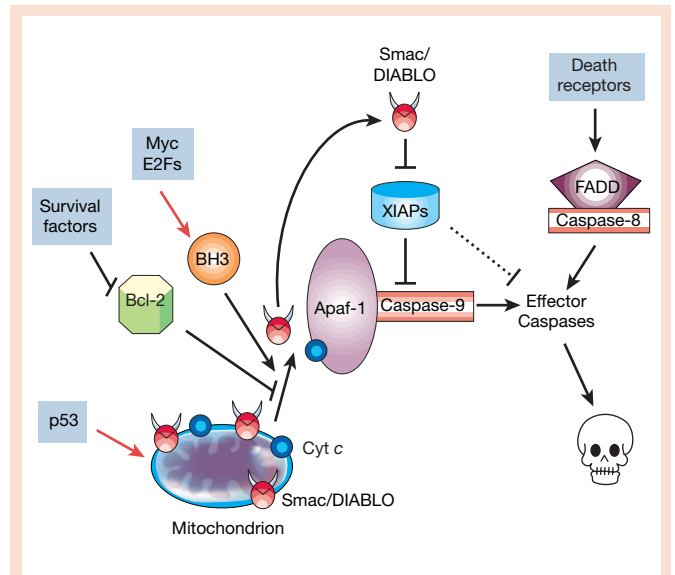


Figure 3 The apoptotic system in mammals. Compare with the apoptotic system in *C. elegans* (Fig. 1) and *D. melanogaster* (Fig. 2). Red arrows indicate possible interactions between components of the apoptotic pathway. Prototypic molecules are used to represent families of proteins; for example, BH3 and Bcl-2 represent pro- and anti-apoptotic members of the Bcl-2 family, respectively.

mis-programming. Perhaps the most patent examples of such somatic confinement are epithelial and endothelial cells that spontaneously commit suicide when detached from their neighbours and basal stroma because they are denied necessary integrin- and cadherin-mediated survival signals. Detachment-induced apoptosis, often termed anoikis, is an important constructive mechanism during development and triggers death in interior cells sundered from outlying basement membranes. Similar cell death also seems necessary during folding, pinching off and fusion of epithelial sheets to generate structures like lens vesicle and vertebrate neural tube. If explanted chick embryos are treated with apoptosis inhibitors the two neural folds still meet but fail to fuse to form the neural tube⁵³.

The profound complexities of the nervous and immune systems make them peculiarly sensitive indicators of perturbation or dysfunction in apoptosis. Perhaps it is no surprise, therefore, that most phenotypes arising from spontaneous or induced mutations in cell death machinery are most evident in these two tissues. Thus, mice lacking caspase-9, caspase-3 or Apaf-1 all exhibit gross neuronal hyperproliferation and disordering^{54–58}, whereas mice deficient in *bcl-x* (encoding both Bcl-x_L and Bcl-x_S) show a marked increase in neuronal apoptosis, leading to embryonic death^{59,60}. Pro-apoptotic signals also are important in neuronal development: for example, trophic withdrawal-induced apoptosis of spinal motor neurons is inhibited by a blocking anti-CD95 antibody⁶¹. Similarly, a host of immunological and haematopoietic phenotypes, some subtle and some dramatic, arise from mutations in genes that regulate or implement apoptosis. The intricate relationships between apoptosis and the nervous and immune systems are dealt with in detail in the accompanying reviews by Yuan and Yankner (pages 802–809) and Krammer (pages 789–795), respectively.

Redundancy in vertebrate cell death

In addition to their complexity, both longevity and size impose additional requirements on the vertebrate cell death machinery. Once crafted, *C. elegans* and *D. melanogaster* have highly restricted regenerative capacities (although this is not always true of other invertebrates). In contrast, apoptosis is required for tissue repair and remodelling throughout vertebrate life in order to cope with the vicissitudes of

infection and injury. Unfortunately, this regenerative capacity of vertebrate tissues carries with it the risk that somatic cells will acquire mutations that confer growth independence and lead to neoplasia. Furthermore, the size and the longevity of vertebrates both conspire to increase the likelihood of such mutations occurring. It is likely that the need to have effective, overlapping and redundant mechanisms to restrict the clonal autonomy of somatic cells has been one of the great imperatives of vertebrate evolution and may well have driven the remarkable redundancy in vertebrate cell death mechanisms.

The dependency of vertebrates on cell death for their greater complexity, plasticity and longevity is reflected in a far more explicit redundancy in mechanisms that regulate cell death than is apparent in invertebrates. The consequences of apoptotic dysfunction on the immune and nervous systems tend to obscure a fact of equally significant biological importance: namely, that most mutations that compromise vertebrate apoptotic machinery have surprisingly little effect on the general abilities of vertebrate tissues to develop. Although many lesions in apoptotic genes are lethal either during embryogenesis or neonatally, death typically results from focal failure in specific tissues and in no case is all embryonic apoptosis blocked. Thus, inactivation of differing caspases induces defects in specific tissues^{62,63}; for example, ablation of caspase-3 and -9 affects brain development whereas loss of caspase-8 affects heart. Likewise, the principle evident effects of loss of TNF and CD95 receptors or ligands are on the immune system⁶⁴. Inactivation of Apaf-1, the ubiquitously expressed adaptor molecule coupling the death pathway to downstream caspases, leads to significant (but not universal) late embryonic death but defects are restricted mainly to brain and craniofacial development and to sterility in surviving males^{57,65}. Indeed, death of the interdigital webs, perhaps the most classical paradigm of morphogenic apoptosis, proceeds unabated in Apaf-1-knockout mice. Even mice lacking cytochrome *c* survive to mid-gestation, by which time a substantial degree of apoptosis-requiring morphogenesis has already occurred.

In part, the robustness of vertebrate cell death reflects an extensive evolutionary duplication and elaboration of apoptotic machinery. At least 15 vertebrate caspases have been identified, four of which seem to be effector caspases whereas the others bear the elaborate prodomains of initiator caspases and are presumably coupled to various upstream pro-apoptotic signals. The vertebrate Bcl-2/BH3 protein family numbers some 20 current members, many of which come in 'various flavours of splice'. Multiple death receptors and cognate ligands have been identified and the number and diversity of signalling pathways that can regulate cell survival seem to be legion. The robustness of vertebrate cell death is probably also indicative of redundancy within the various cellular pathways that conspire to create the apoptotic process. Evidence indicates that 'apoptosis' comprises at least three discrete, if intertwined, mechanisms, any one of which would alone be sufficient for cellular demise. In addition to caspase activation, most pro-apoptotic insults also trigger mitochondrial dysfunction⁶⁶ and expression of pro-phagocytic signals (see review in this issue by Savill and Fadok, pages 784–788), neither of which necessarily depends upon caspase activity. Consistent with this, there are reported instances of inhibition of caspase activation re-routing cells into abortive or necrotic forms of cell death, but cell death nonetheless^{67–69}. More generally still, much physiological cell death in vertebrates may be triggered by non-apoptotic mechanisms. For example, in superficial epithelia such as skin and gastrointestinal tract, arguably the tissues at most risk of carcinogenic insult and neoplastic mutation, the inescapable death of progeny cells is guaranteed by a combination of irreversible post-mitotic terminal differentiation and the simple expedient of detachment and shedding. Whether such detached cells die by apoptosis or necrosis is unclear. However, the distinction is immaterial since suppression of apoptosis, for example by transgenic expression of Bcl-2 or Bcl-x_L, has no observable effect on terminal differentiation and cell loss in either skin⁷⁰ or intestinal epithelium⁷¹.

To cope with the relentless risk of cancer, vertebrates also use cell death as an adaptive mechanism to ablate rogue or neoplastic cells. One way this is achieved is through the tight coupling of cell proliferative and apoptotic programmes such that all cells forced into a proliferative state, and therefore a potential neoplastic risk to the host, are rendered acutely sensitive to induction of apoptosis (reviewed in ref. 72). The molecular basis of this coupling seems to involve at least three independent mechanisms. First, oncoproteins like Myc or the E2F G1-progression transcription factors are potent inducers of release of cytochrome *c* from mitochondria, which can trigger activation of the Apaf-1/caspase-9 apoptotic cascade⁷³. Such cytochrome *c* release is suppressed by survival signals, ensuring that activation of growth-promoting oncogenes triggers apoptosis should the affected cell or its progeny stray out of their orthodox trophic environment. Second, growth-promoting oncoproteins induce expression of p53 (ref. 74). This induces a state of extreme sensitivity to DNA damage or cellular stress, upon which the affected cell either arrests or commits suicide (see review in this issue by Rich *et al.*, pages 777–783). Third, expression of many oncoproteins induces rapid downregulation of cadherins, triggering a *defacto* state of anoikis unless the affected cell can expeditiously re-establish appropriate attachments.

An understanding of the mechanisms controlling and implementing apoptosis is more than a matter of mere scientific interest. Apoptosis is an essential component of most developmental abnormalities and human diseases and, in many cases, the underlying cause of the resulting pathology. Disorders associated with insufficient cell death include autoimmunity and cancer, but it has also become clear that many, if not all, viruses possess mechanisms to forestall apoptosis and provide a living host to nurture virus propagation⁷⁵. In such cases, reinstating the blocked or defective apoptotic programme is likely to have an enormous impact on the disease. On the other hand, many other diseases including AIDS, stroke and neurodegenerative disorders such as Alzheimer's, Parkinson's and retinitis pigmentosa involve excessive apoptosis. In such instances, suppression of apoptosis may restore functionality to the affected tissue (see review in this issue by Nicholson, pages 810–816). The conservation of apoptotic machinery through evolution has provided us with a wealth of experimental systems with which to study, understand and, eventually, manipulate this fundamental biological process. □

- Jacobson, M. D., Weil, M. & Raff, M. C. Programmed cell death in animal development. *Cell* **88**, 347–354 (1997).
- Saxén, L. *Organogenesis of the Kidney* (Cambridge Univ. Press, Cambridge, 1987).
- Namba, R., Pazdera, T. M., Cerrone, R. L. & Minden, J. S. *Drosophila* embryonic pattern repair: how embryos respond to bicoid dosage alteration. *Development* **124**, 1393–1403 (1997).
- Thornberry, N. A. & Lazebnik, Y. Caspases: enemies within. *Science* **281**, 1312–1316 (1998).
- Hengartner, M. Apoptosis. Death by crowd control. *Science* **281**, 1298–1299 (1998).
- Shaham, S. Identification of multiple *Caenorhabditis elegans* caspases and their potential roles in proteolytic cascades. *J. Biol. Chem.* **273**, 35109–35117 (1998).
- Conradt, B. & Horvitz, H. R. The TRA-1A sex determination protein of *C. elegans* regulates sexually dimorphic cell deaths by repressing the *egl-1* cell death activator gene. *Cell* **98**, 317–327 (1999).
- Inukai, T. *et al.* SLUG, a CES-1-related zinc finger transcription factor gene with antiapoptotic activity, is a downstream target of the E2A-HLF oncoprotein. *Mol. Cell* **4**, 343–352 (1999).
- Metzstein, M. M. & Horvitz, H. R. The *C. elegans* cell death specification gene *ces-1* encodes a SNAIL family zinc finger protein. *Mol. Cell* **4**, 309–319 (1999).
- Inaba, T. *et al.* Reversal of apoptosis by the leukaemia-associated E2A-HLF chimaeric transcription factor. *Nature* **382**, 541–544 (1996).
- Metzstein, M. M., Hengartner, M. O., Tsung, N., Ellis, R. E. & Horvitz, H. R. Transcriptional regulator of programmed cell death encoded by *Caenorhabditis elegans* gene *ces-2*. *Nature* **382**, 545–547 (1996).
- Gartner, A., Milstein, S., Ahmed, S., Hodgkin, J. & Hengartner, M. O. A conserved checkpoint pathway mediates DNA damage-induced apoptosis and cell cycle arrest in *C. elegans*. *Mol. Cell* **5**, 435–443 (2000).
- Gumienny, T. L., Lambie, E., Hartwig, E., Horvitz, H. R. & Hengartner, M. O. Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline. *Development* **126**, 1011–1022 (1999).
- Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome *c*-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
- Kanuka, H. *et al.* Control of the cell death pathway by Dapaf-1, a *Drosophila* Apaf-1/CED-4-related caspase activator. *Mol. Cell* **4**, 757–769 (1999).
- Zhou, L., Song, Z., Tittel, J. & Steller, H. HAC-1, a *Drosophila* homolog of APAF-1 and CED-4 functions in developmental and radiation-induced apoptosis. *Mol. Cell* **4**, 745–755 (1999).
- Rodriguez, A. *et al.* Dark is a *Drosophila* homolog of Apaf-1/CED-4 and functions in an

- evolutionarily conserved death pathway. *Nature Cell Biol.* **1**, 272–279 (1999).
18. Antonsson, B. & Martinou, J. C. The Bcl-2 protein family. *Exp. Cell Res.* **256**, 50–57 (2000).
19. Brachmann, C. B., Jassim, O. W., Wachsmuth, B. D. & Cagan, R. L. The *Drosophila* Bcl-2 family member dBorg-1 functions in the apoptotic response to UV-irradiation. *Curr. Biol.* **10**, 547–550 (2000).
20. Igaki, T. *et al.* Drob-1, a *Drosophila* member of the Bcl-2/CED-9 family that promotes cell death. *Proc. Natl Acad. Sci. USA* **97**, 662–667 (2000).
21. Colussi, P. A. *et al.* Debcl, a proapoptotic Bcl-2 homologue, is a component of the *Drosophila melanogaster* cell death machinery. *J. Cell Biol.* **148**, 703–714 (2000).
22. Zhang, H. *et al.* *Drosophila* Pro-apoptotic Bcl-2/Bax homologue reveals evolutionary conservation of cell death mechanisms. *J. Biol. Chem.* **275**, 27303–27306 (2000).
23. Kelekar, A. & Thompson, C. B. Bcl-2-family proteins: the role of the BH3 domain in apoptosis. *Trends Cell Biol.* **8**, 324–330 (1998).
24. Vaux, D. L., Weissman, I. L. & Kim, S. K. Prevention of programmed cell-death in *Caenorhabditis elegans* by human bcl-2. *Science* **258**, 1955–1957 (1992).
25. White, K. *et al.* Genetic control of programmed cell death in *Drosophila*. *Science* **264**, 677–683 (1994).
26. Ollmann, M. *et al.* *Drosophila* p53 is a structural and functional homolog of the tumor suppressor p53. *Cell* **101**, 91–101 (2000).
27. Jin, S. *et al.* Identification and characterization of a p53 homologue in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 7301–7306 (2000).
28. Brodsky, M. H. *et al.* *Drosophila* p53 binds a damage response element at the *reaper* locus. *Cell* **101**, 103–113 (2000).
29. Bergmann, A., Agapite, J., McCall, K. & Steller, H. The *Drosophila* gene *hid* is a direct molecular target of Ras-dependent survival signaling. *Cell* **95**, 331–341 (1998).
30. Kurada, P. & White, K. Ras promotes cell survival in *Drosophila* by down-regulating *hid* expression. *Cell* **95**, 319–329 (1998).
31. Foley, K. & Cooley, L. Apoptosis in late stage *Drosophila* nurse cells does not require genes within the *H99* deficiency. *Development* **125**, 1075–1082 (1998).
32. Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
33. Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
34. Chai, J., Du, C., Wu, J.-W., Wang, X. & Shi, Y. Structural and biochemical basis of apoptotic activation by Smac/DIABLO. *Nature* **406**, 855–862 (2000).
35. Goyal, L., McCall, K., Agapite, J., Hartwig, E. & Steller, H. Induction of apoptosis by *Drosophila reaper*, *hid* and *grim* through inhibition of IAP function. *EMBO J.* **19**, 589–597 (2000).
36. Meier, P., Silke, J., Leivers, S. J. & Evan, G. I. The *Drosophila* caspase DRONC is regulated by DIAP1. *EMBO J.* **19**, 598–611 (2000).
37. Song, Z. *et al.* Biochemical and genetic interactions between *Drosophila* caspases and the proapoptotic genes *rpr*, *hid*, and *grim*. *Mol. Cell Biol.* **20**, 2907–2914 (2000).
38. Wang, S. L., Hawkins, C. J., Yoo, S. J., Muller, H. A. & Hay, B. A. The *Drosophila* caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID. *Cell* **98**, 453–463 (1999).
39. Chen, P. & Abrams, J. M. *Drosophila* apoptosis and Bcl-2 genes: outliers fly in. *J. Cell Biol.* **148**, 625–627 (2000).
40. Bergmann, A., Agapite, J. & Steller, H. Mechanisms and control of programmed cell death in invertebrates. *Oncogene* **17**, 3215–3223 (1998).
41. McCall, K. & Steller, H. Requirement for DCP-1 caspase during *Drosophila* oogenesis. *Science* **279**, 230–234 (1998).
42. Jiang, C., Baehrecke, E. H. & Thummel, C. S. Steroid regulated programmed cell death during *Drosophila* metamorphosis. *Development* **124**, 4673–4683 (1997).
43. Tata, J. R. Requirement for RNA and protein synthesis for induced regression of the tadpole tail in organ culture. *Dev. Biol.* **13**, 77–94 (1966).
44. Kiess, W. & Gallaher, B. Hormonal control of programmed cell death/apoptosis. *Eur. J. Endocrinol.* **138**, 482–491 (1998).
45. Merino, R., Ganan, Y., Macias, D., Rodriguez-Leon, J. & Hurlé, J. M. Bone morphogenetic proteins regulate interdigital cell death in the avian embryo. *Ann. NY Acad. Sci.* **887**, 120–132 (1999).
46. Hu, S. & Yang, X. dFADD, a novel death domain-containing adapter protein for the *Drosophila* caspase DREDD. *J. Biol. Chem.* (in the press).
47. Kondo, T., Yokokura, T. & Nagata, S. Activation of distinct caspase-like proteases by Fas and reaper in *Drosophila* cells. *Proc. Natl Acad. Sci. USA* **94**, 11951–11956 (1997).
48. Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1-Cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
49. Varkey, J., Chen, P., Jemmerson, R. & Abrams, J. M. Altered cytochrome c display precedes apoptotic cell death in *Drosophila*. *J. Cell Biol.* **144**, 701–710 (1999).
50. Barres, B. A. & Raff, M. C. Axonal control of oligodendrocyte development. *J. Cell Biol.* **147**, 1123–1128 (1999).
51. Raff, M. C. Social controls on cell survival and cell death. *Nature* **356**, 397–400 (1992).
52. Raff, M. C. *et al.* Programmed cell death and the control of cell survival: lessons from the nervous system. *Science* **262**, 695–700 (1993).
53. Weil, M., Jacobson, M. D. & Raff, M. C. Is programmed cell death required for neural tube closure? *Curr Biol* **7**, 281–284 (1997).
54. Kuida, K. *et al.* Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* **384**, 368–372 (1996).
55. Kuida, K. *et al.* Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* **94**, 325–337 (1998).
56. Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell* **94**, 339–352 (1998).
57. Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
58. Cecconi, F., Alvarez-Bolado, G., Meyer, B. I., Roth, K. A. & Gruss, P. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
59. Middleton, G., Cox, S. W., Korsmeyer, S. & Davies, A. M. Differences in Bcl-2- and Bax-independent function in regulating apoptosis in sensory neuron populations. *Eur. J. Neurosci.* **12**, 819–827 (2000).
60. Motoyama, N. *et al.* Massive cell death of immature hematopoietic cells and neurons in Bcl-x-deficient mice. *Science* **267**, 1506–1510 (1995).
61. Raoul, C., Henderson, C. E. & Pettmann, B. Programmed cell death of embryonic motoneurons triggered through the Fas death receptor. *J. Cell Biol.* **147**, 1049–1062 (1999).
62. Zheng, T. S., Hunot, S., Kuida, K. & Flavell, R. A. Caspase knockouts: matters of life and death. *Cell Death Differ.* **6**, 1043–1053 (1999).
63. Wang, J. & Lenardo, M. J. Roles of caspases in apoptosis, development, and cytokine maturation revealed by homozygous gene deficiencies. *J. Cell Sci.* **113**, 753–757 (2000).
64. Yeh, W. C., Hakem, R., Woo, M. & Mak, T. W. Gene targeting in the analysis of mammalian apoptosis and TNF receptor superfamily signaling. *Immunol. Rev.* **169**, 283–302 (1999).
65. Honarpour, N. *et al.* Adult Apaf-1-deficient mice exhibit male infertility. *Dev. Biol.* **218**, 248–258 (2000).
66. Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
67. Xiang, J., Chao, D. & Korsmeyer, S. Bax-induced cell death may not require interleukin-1 β -converting enzyme-like proteases. *Proc. Natl Acad. Sci. USA* **93**, 14559–14563 (1996).
68. McCarthy, N., Whyte, M., Gilbert, C. & Evan, G. Inhibition of Ced-3/ICE-related proteases does not prevent cell death induced by oncogenes, DNA damage, or the Bcl-2 homologue Bak. *J. Cell Biol.* **136**, 215–227 (1997).
69. Chautan, M., Chazal, G., Cecconi, F., Gruss, P. & Golstein, P. Interdigital cell death can occur through a necrotic and caspase-independent pathway. *Curr. Biol.* **9**, 967–970 (1999).
70. Pena, J. C., Fuchs, E. & Thompson, C. B. Bcl-x expression influences keratinocyte cell survival but not terminal differentiation. *Cell Growth Differ.* **8**, 619–629 (1997).
71. Merriitt, A. J. *et al.* Differential expression of *bcl-2* in intestinal epithelia. Correlation with attenuation of apoptosis in colonic crypts and the incidence of colonic neoplasia. *J. Cell Sci.* **108**, 2261–2271 (1995).
72. Evan, G. & Littlewood, T. A matter of life and cell death. *Science* **281**, 1317–1322 (1998).
73. Juin, P., Hueber, A. O., Littlewood, T. & Evan, G. c-Myc-induced sensitization to apoptosis is mediated through cytochrome c release. *Genes Dev.* **13**, 1367–1381 (1999).
74. Sherr, C. J. & Weber, J. D. The ARF/p53 pathway. *Curr. Opin. Genet. Dev.* **10**, 94–99 (2000).
75. Roulston, A., Marcellus, R. C. & Branton, P. E. Viruses and apoptosis. *Annu. Rev. Microbiol.* **53**, 577–628 (1999).

Apoptosis in the nervous system

Junying Yuan* & Bruce A. Yankner†

*Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA (email: junying_yuan@hms.harvard.edu)

†Department of Neurology, Harvard Medical School and Division of Neuroscience, Children's Hospital, Enders 260, 300 Longwood Avenue, Boston, Massachusetts 02115, USA (e-mail: yankner@a1.tch.harvard.edu)

Neuronal apoptosis sculpts the developing brain and has a potentially important role in neurodegenerative diseases. The principal molecular components of the apoptosis programme in neurons include Apaf-1 (apoptotic protease-activating factor 1) and proteins of the Bcl-2 and caspase families. Neurotrophins regulate neuronal apoptosis through the action of critical protein kinase cascades, such as the phosphoinositide 3-kinase/Akt and mitogen-activated protein kinase pathways. Similar cell-death-signalling pathways might be activated in neurodegenerative diseases by abnormal protein structures, such as amyloid fibrils in Alzheimer's disease. Elucidation of the cell death machinery in neurons promises to provide multiple points of therapeutic intervention in neurodegenerative diseases.

Although mature neurons are among the most long-lived cell types in mammals, immature neurons die in large numbers during development. Furthermore, neuronal cell death is the cardinal feature of both acute and chronic neurodegenerative diseases. How do neurons die? This is a difficult question and we have only recently begun to understand the basic mechanisms. Like all cells, neuronal survival requires trophic support. Viktor Hamburger and Rita Levi-Montalcini described in a seminal paper that the survival of developing neurons is directly related to the availability of their innervating targets¹. This laid the foundation for the neurotrophin hypothesis², which proposed that immature neurons compete for target-derived trophic factors that are in limited supply; only those neurons that are successful in establishing correct synaptic connections would obtain trophic factor support to allow their survival. The neurotrophin hypothesis predicts correctly that neuronal survival requires a positive survival signal; it did not, however, provide a concrete hypothesis as to how neurons die in the absence of trophic support.

It was assumed until recently that neurons die simply of passive starvation in the absence of trophic factors. In 1988, using cultured sympathetic neurons as a model system, Johnson and colleagues showed that inhibition of RNA and protein synthesis blocked sympathetic neuronal cell death induced by nerve growth factor (NGF) deprivation³, providing the first tangible evidence that neurons might actually instigate their own demise. The identification of the programmed cell death genes *ced-3*, *ced-4* and *ced-9*, in the nematode *Caenorhabditis elegans* and their mammalian homologues (see review in this issue by Meier *et al.*, pages 796–801) opened a window of opportunity to examine the mechanism of neuronal cell death at the molecular level⁴. It was soon discovered that vertebrate neuronal cell death induced by trophic factor deprivation requires the participation of cysteine proteases, later termed caspases, which are the mammalian homologues of the *C. elegans* cell death gene product CED-3 (ref. 5). This was the first functional evidence that trophic factor deprivation activates a cellular suicide programme in vertebrate neurons. What are the critical components of this neuronal suicide programme? How is it activated by lack of trophic support during

development and by pathological conditions in neurodegenerative diseases? These questions have been studied intensively during the past decade and are the subject of this review.

Key molecules in neuronal apoptosis

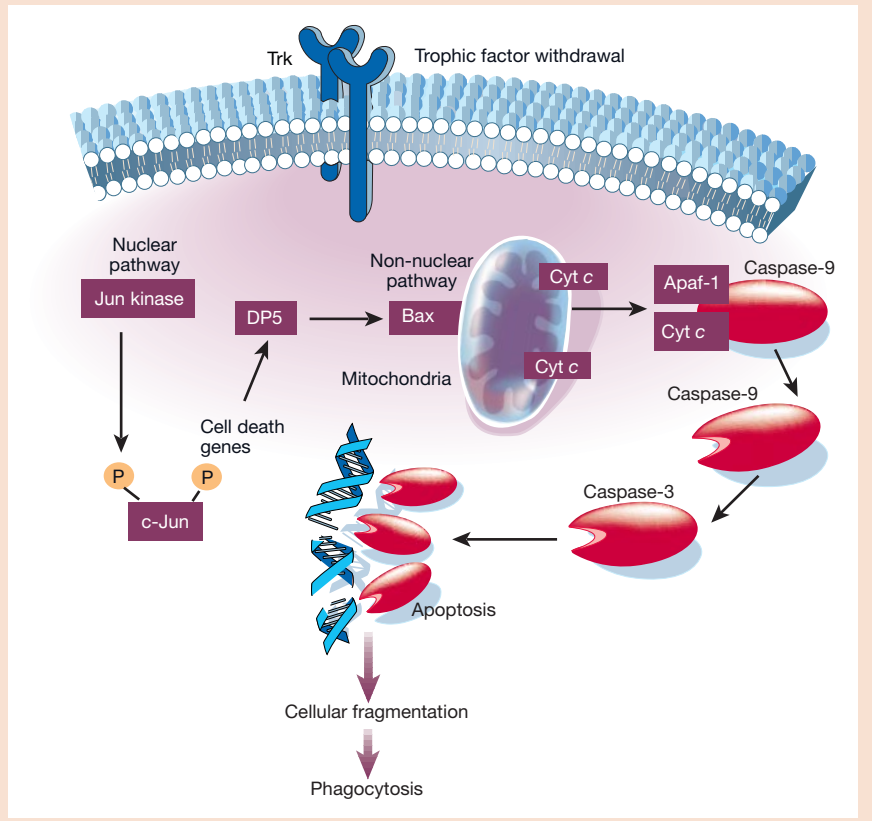
Mammalian apoptosis is regulated by the Bcl-2 family of proteins, the adaptor protein Apaf-1 (for apoptotic protease-activating factor 1) and the cysteine protease caspase family, which are homologues of the *C. elegans* cell-death gene products CED-9, CED-4 and CED-3, respectively (see review in this issue by Hengartner, pages 770–776). Neurons share the same basic apoptosis programme with all other cell types. However, different types of neurons, and neurons at different developmental stages, express different combinations of Bcl-2 and caspase family members, which is one way of providing the specificity of regulation.

The role of the Bcl-2 family in neuronal cell death

The Bcl-2 family of proteins has a crucial role in intracellular apoptotic signal transduction. This gene family includes both anti-apoptotic and pro-apoptotic proteins that contain one or more Bcl-2 homology (BH) domains⁶. The major anti-apoptotic members of the Bcl-2 family, Bcl-2 and Bcl-x_L, are localized to the mitochondrial outer membrane and to the endoplasmic reticulum and perinuclear membrane. Garcia *et al.*⁷ showed that Bcl-2 can support the survival of sympathetic neurons in the absence of NGF, providing the first functional evidence that the overexpression of Bcl-2 can override the death signal induced by the withdrawal of a trophic factor. Subsequently, transgenic mice expressing Bcl-2 in the nervous system were found to be protected against neuronal cell death during development⁸, as were neuronal injury models such as middle cerebral artery occlusion and facial nerve axotomy^{9,10}. These results suggest that the suppression of apoptosis might protect neurons against insults ranging from trophic factor deprivation to pathological stimuli.

The expression of Bcl-2 is high in the central nervous system during development and is downregulated after birth, whereas the expression of Bcl-2 in the peripheral nervous system is maintained throughout life⁶. Although the development of the nervous system in Bcl-2-knockout mice is normal, there is a subsequent loss of motor, sensory and sympathetic neurons after birth^{11,12}, suggesting that Bcl-2 is crucial for the maintenance of neuronal survival. Bcl-x_L is

Figure 1 Activation of apoptosis in sympathetic neurons by trophic factor withdrawal. Trophic factor withdrawal induces JNK activation and the phosphorylation of c-Jun, which in turn induces the expression of DP5/Hrk, a 'BH3-domain only' member of the Bcl-2 family. DP5 might activate Bax, causing mitochondrial damage, which results in the release of cytochrome *c*. Formation of the cytochrome *c*/Apaf-1/caspase-9 complex induces the activation of caspase-9. Activated caspase-9 in turn activates caspase-3, resulting in apoptosis. A lack of trophic factor signalling also induces a non-nuclear competence-to-die pathway that facilitates the formation of the cytochrome *c*/Apaf-1/caspase-9 complex, resulting in caspase-9 activation.



expressed in developing brain; but unlike Bcl-2 expression, Bcl-x_L expression continues to increase into adult life¹³. Bcl-x_L-null mice die around embryonic day 13 with massive cell death in the developing nervous system¹⁴. Cell death occurs primarily in immature neurons that have not established synaptic connections. Thus, Bcl-x_L might be critical for the survival of immature neurons before they establish synaptic connections with their targets.

Bcl-2 and Bcl-x_L act by inhibiting pro-apoptotic members of the Bcl-2 family through heterodimerization⁶. Bax is a pro-apoptotic member of the Bcl-2 family that is widely expressed in the nervous system¹⁵. In Bax-deficient mice, superior cervical ganglia and facial nuclei display increased neuron number. Furthermore, neonatal sympathetic neurons and facial motor neurons from Bax-deficient mice are more resistant to cell death induced by NGF deprivation and axotomy, respectively. Thus, the activation of Bax might be a crucial event for neuronal cell death induced by trophic factor withdrawal as well as injury.

Apaf-1 and caspases in neuronal cell death

Apaf-1 is a mammalian homologue of the *C. elegans* cell-death gene product CED-4 and transmits apoptotic signals from mitochondrial damage to activate caspases. Apaf-1 forms a complex with mitochondrial-released cytochrome *c* and caspase-9 to mediate the activation of pro-caspase-9 (see Fig. 1)¹⁶. Activated caspase-9 in turn cleaves and activates caspase-3. Apaf-1-null mice die during late embryonic development, exhibiting reduced apoptosis in the brain with a marked enlargement of the periventricular proliferative zone¹⁷. Thus, Apaf-1 is indispensable in the apoptosis of neuronal progenitor cells.

The ability of caspase inhibitors to block neuronal cell death induced by trophic factor deprivation and other cytotoxic conditions has provided indisputable evidence for a crucial role of caspases in neuronal cell death¹⁸. But it has been more challenging to determine the role of specific caspases because mammals have at least 14 different caspases. Like other cell types, neurons can express several of them simultaneously. This has ruled out the simplistic model that neuronal cell death is regulated by neuron-specific caspases. Instead, biochemical and genetic analysis of caspase-mutant mice suggest

that caspases are organized into parallel and sometimes overlapping pathways that are specialized to respond to different stimuli. Caspases are expressed as catalytically inactive proenzymes composed of an amino-terminal pro-domain, a large subunit and a small subunit. Caspases can be classified on the basis of the sequence motifs in their pro-domains. Caspases with the death-effector domain, which include caspase-8 and caspase-10, are activated by interacting with the intracellular domains of death receptors such as the CD95 (Apo-1/Fas) and tumour necrosis factor (TNF) receptors. Caspases with caspase-activating recruitment domains (CARDs), which include caspase-1, -2, -4, -5, -9, -11 and -12, are most probably activated through an intracellular activating complex exemplified by the cytochrome *c*/Apaf-1/caspase-9 complex¹⁹. Whereas caspases with short pro-domains, such as caspase-3, might be activated by most, if not all, caspase pathways, recent data indicate that some caspases, such as caspase-11 and caspase-12, are activated only under pathological conditions^{20,21}. This offers the prospect of being able to inhibit pathological cell death therapeutically without disturbing developmental and homeostatic apoptosis (see review in this issue by Nicholson, pages 810–816).

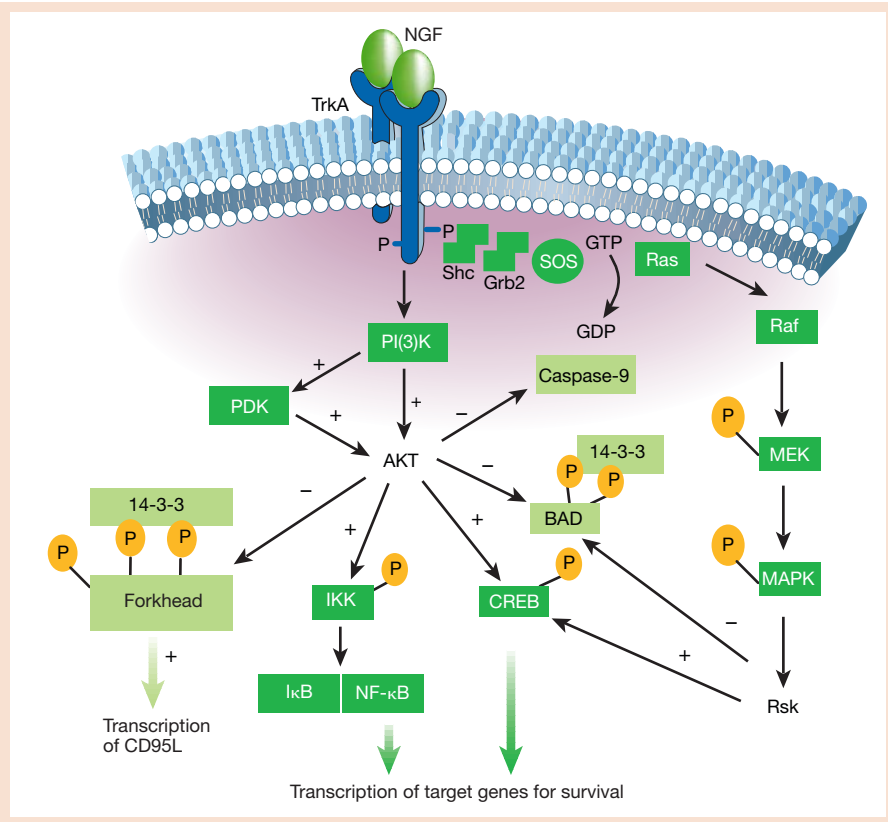
The two major caspases involved in neuronal cell death are caspase-3 and caspase-9, in which the latter activates the former (Fig. 1). Both caspase-3-null²² and caspase-9-null²³ mice show severe and similar defects in developmental neuronal cell death. Ectopic cell masses appear in the cerebral cortex, hippocampus and striatum of the caspase-3-null and caspase-9-null mice with marked expansion of the periventricular zone, a phenotype very similar to that of Apaf-1-null mice¹⁷. The prominent neuronal apoptosis defects of Apaf-1-null, caspase-3-null and caspase-9-null mice (Table 1) suggest that this pathway is important in regulating neuronal cell death in the developing brain.

Neurotrophins: a matter of life and death

Staying alive with neurotrophins

As mentioned above, the survival of developing immature neurons depends on the availability of neurotrophic factors. What do these

Figure 2 Neuronal survival pathways induced by the binding of NGF to its receptor TrkA. NGF induces the autophosphorylation of TrkA which provides docking sites for signal transduction molecules such as phospholipase C γ , phosphoinositide 3-kinase (PI(3)K) and the adaptor protein Shc. Activated PI(3)K induces the activation of Akt through 3'-phosphorylated phosphatidylinositol as well as phosphoinositide-dependent kinase (PDK), which in turn phosphorylates and activates Akt. The phosphorylation of CREB and IKK stimulates the transcription of pro-survival factors; whereas the phosphorylation of Bad, Forkhead and caspase-9 inhibits the pro-apoptotic pathway. In a parallel pathway, the interaction of Shc-Grb2 and SOS activates the Ras-Raf-MEK-ERK pathway, resulting in the activation of Rsk. Bad and CREB are also the targets of Rsk that might act synergistically with Akt to activate the survival pathway.



survival factors do? Neurotrophins generally activate and ligate the Trk receptors (TrkA, TrkB and TrkC), which are cell-surface receptors with intrinsic tyrosine kinase activity. They can autophosphorylate²⁴; for instance, after the binding of NGF to TrkA, the receptor phosphorylates several tyrosine residues within its own cytoplasmic tail. These phosphotyrosines in turn serve as docking sites for other molecules such as phospholipase C γ , phosphoinositide 3-kinase (PI(3)K)²⁵ and adaptor proteins such as Shc, and these signal transduction molecules coordinate neuronal survival (Fig. 2).

PI(3)K-Akt pathway. A central role of the PI(3)K pathway in neuronal survival was first suggested by the observation that PI(3)K inhibitors block the survival effect of NGF²⁶. PI(3)K enzymes are normally present in cytosol and can be activated directly by recruitment to an activated Trk receptor, or indirectly through activated Ras. Active PI(3)K enzymes catalyse the formation of the lipid 3'-phosphorylated phosphoinositides, which regulate the localization and activity of a key component in cell survival, the Ser/Thr kinase Akt (ref. 27).

Akt has three cellular isoforms, of which c-Akt3/RAC-PK γ is the major species expressed in neurons²⁸. In addition to a centrally located kinase domain, Akt contains a pleckstrin homology domain at its N-terminus, which mediates its interaction with proteins and phospholipids. After the binding of lipid, Akt is translocated from the cytoplasm to the inner surface of the plasma membrane, which brings the kinase into close proximity with its activators. The kinases that phosphorylate and activate Akt, the 3-phosphoinositide-dependent protein kinases are — as their name suggests — themselves regulated by phospholipids. Thus, the lipid products generated by PI(3)K enzymes control the activity of Akt by regulating its location and activation.

Active Akt protein supports the survival of neurons in the absence of trophic factors, whereas a dominant-negative mutant of Akt inhibits neuronal survival even in the presence of survival factors²⁸. These results establish an essential role for Akt in neuronal survival. How does Akt act?

Akt in action. Akt targets several key proteins to keep cells alive, including apoptosis regulators and transcription factors (Fig. 2). For example,

Bad is a pro-apoptotic member of the Bcl-2 family, which in its unphosphorylated form can bind to Bcl-x_L and thus block cell survival²⁹. But the activation of Akt induces the phosphorylation of Bad and promotes its interaction with the chaperone protein 14-3-3, which sequesters Bad in the cytoplasm and inhibits Bad's pro-apoptotic activity³⁰. Akt has been shown to affect, directly or indirectly, three transcription factor families: Forkhead, cAMP-response-element-binding protein (CREB) and NF- κ B, all of which are involved in regulating cell survival, and whereas the phosphorylation of Forkhead family members by Akt negatively regulates death-promoting signals³¹, the phosphorylation of CREB and I κ B kinase (IKK) stimulates survival pathways³²⁻³⁴. It is clear that Akt is a potent kinase that keeps neurons alive in various ways, and that additional targets of Akt will no doubt be identified.

Mitogen-activated protein (MAP) kinase pathway. But there is more to neurotrophins than only the activation of PI(3)K and Akt: they also stimulate docking of the adaptor protein Shc to activated Trk receptors. This triggers the activation of the small GTP-binding protein Ras and the downstream MAP kinase cascade, which includes the subsequent sequential phosphorylation and activation of the kinases Raf, MAP kinase/ERK kinase (MEK) and extracellular signal-regulated protein kinase (ERK)³⁵ (Fig. 2). The effect of the MAP kinase pathway on survival is mediated at least partly by activation of the pp90 ribosomal S6 kinase (RSK) family members. Like Akt, RSK phosphorylates Bad, and both kinases might act synergistically in inhibiting Bad's pro-apoptotic activity. The effect of RSKs on neuronal survival is not limited to the phosphorylation of Bad; RSKs are also potent activators of the CREB transcription factor. Because CREB is known to activate transcription of *bcl-2*, it can stimulate cell survival directly. Thus, although there is a divergence in the survival pathways downstream of the neurotrophin receptors, both the PI(3)K-Akt and MAP kinase pathways converge on the same set of proteins, Bad and CREB, to inhibit the apoptosis programme.

It is noteworthy that neurotrophins are not the only factors that promote neuronal survival: electrical stimulation and depolarization at high KCl concentration have long been known to inhibit neuronal cell

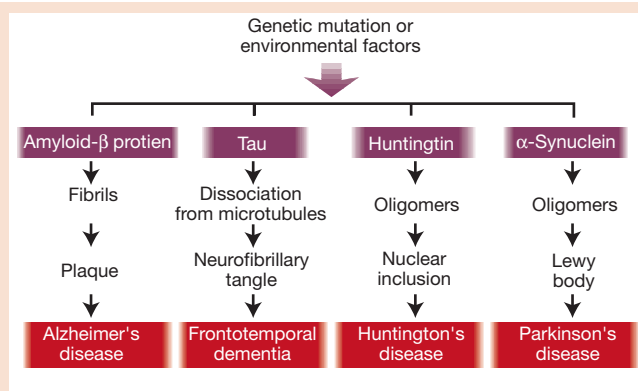


Figure 3 Abnormal protein structures and the pathogenesis of neurodegenerative disease. Normal proteins might become pathogenic when subjected to genetic mutations or environmental factors that promote the formation of abnormal structures in specific neuronal subpopulations.

death³⁶. Recent studies indicate that membrane depolarization also activates neuronal survival pathways; whether or not these are the same as those activated by the neurotrophins is unresolved^{37,38}.

Dying without neurotrophins

Although it is clear that neurotrophins and membrane depolarization activate signal transduction pathways that suppress apoptosis, it is less clear what triggers the activation of apoptosis in the absence of survival signals. It is possible that neurotrophins simply suppress a default apoptosis programme. However, a number of processes need to happen before cultured immature sympathetic neurons are committed to die (Fig. 1).

The removal of NGF results in a decrease in MAP kinase and PI(3)K activities, followed by a series of early metabolic changes including the increased production of reactive oxygen species, decreased glucose uptake and decreased RNA and protein synthesis. In some cells, the removal of NGF results in a slow and sustained increase in c-Jun amino-terminal kinase (JNK) and p38 MAP kinase activities³⁹; in other cells, c-Jun, one of the downstream targets of JNK, is induced and phosphorylated^{40,41}. The activation of JNK itself might be necessary, but not sufficient, to induce neuronal apoptosis.

Paradoxically, although protein and RNA synthesis are significantly reduced in the early stages of sympathetic neuronal cell death, death cannot occur in the presence of inhibitors of RNA and protein synthesis, indicating that the continued synthesis of certain pro-apoptotic molecules is required. In view of this, it is interesting to note that DP5 (also known as Hrk), a 'BH3-domain only' pro-apoptotic member of the Bcl-2 family, is induced in NGF-deprived sympathetic neurons⁴². Perhaps it is the synthesis of DP5-related proteins that is vital to the execution of the cell death programme. DP5 could be required to help Bax move from its location in the cytosol to the mitochondria, after which Bax can induce the release of cytochrome *c* (ref. 43) (Fig. 1).

As in other cell types, the release of cytochrome *c* from mitochondria

induces the activation of caspases in sympathetic neurons. The addition of a pan-caspase inhibitor, but not NGF, rescues sympathetic neurons even after mitochondrial damage and the release of cytochrome *c*⁴⁴. Thus, these neurons are not committed to die until caspases are fully activated. This indicates that the point of no return is at, or downstream of, caspase activation, and suggests that the inhibition of caspase activity might be sufficient to block neuronal cell death under certain pathological conditions.

Neurotrophins: a double-edged sword?

The neurotrophin hypothesis predicts that neurotrophins function as survival signals to suppress the death programme. However, the interaction of neurotrophins with the neurotrophin receptor p75NTR can induce cell death under certain conditions, suggesting that neurotrophins might act as death ligands in a cell-context-dependent manner. The p75 neurotrophin receptor (p75NTR) is a member of the TNF receptor superfamily that can bind all neurotrophins⁴⁵. Its intracellular domain contains a region that bears similarity with the 'death domain', which mediates protein-protein interactions and is present in other members of the TNF family. p75NTR was originally thought to cooperate with Trks to modulate the response to neurotrophins. However, p75NTR might have an additional role in orchestrating neuronal cell death. Barde and colleagues found that application of antibodies that block the binding of NGF to p75NTR inhibited the death of chick retinal ganglion cells that express p75NTR but not trkA⁴⁶, indicating that the interaction of NGF with p75NTR promotes cell death in this system. Death-inducing activity of various neurotrophins has now been documented for different neuronal cell types^{47,48}. p75NTR-dependent cell death seems to be inhibited by Trk signalling⁴⁹. In view of these opposing roles of neurotrophins, they might be more appropriately referred to as 'neuromodulators' that function to adjust neuronal cell number and regulate differentiation.

Pathological apoptosis in the adult brain

Physiological apoptosis in the developing brain and pathological apoptosis in the adult brain share similar molecular mechanisms in the effector phase. But there are key differences in the mechanisms by which apoptosis is triggered. Whereas trophic factor withdrawal has a prominent role in apoptosis during development, there is little evidence to implicate trophic factor withdrawal as a primary pathogenic mechanism in adult neurodegenerative disorders. Rather, toxic insults resulting from biochemical or genetic accidents might trigger neurodegenerative diseases by co-opting apoptotic signalling pathways, for example through free-radical generation or caspase activation. An emerging theme in adult neurodegenerative disorders is the toxicity of abnormal protein structures or aggregates, which might be important in the pathogenesis of Alzheimer's disease, Parkinson's disease, Huntington's disease and amyotrophic lateral sclerosis (Fig. 3).

Cell death due to ischaemia

Ischaemic injury-induced neuronal cell death has traditionally been characterized as necrosis, in which cells and their organelles swell and rupture. However, morphological and biochemical evidence of apoptosis have now been well documented in experimental animal models of ischaemic brain injury. Apoptotic neurons are more easily

Table 1 Neuronal phenotypes of caspase-knockout mice

Caspase	Development	Neuronal apoptotic phenotype	References
Caspase-1	Normal	Resistant to ischaemic brain injury and a decrease in incidence and severity of experimental autoimmune encephalomyelitis	98
Caspase-2	Normal	Hippocampal neurons from caspase-2-null mice are resistant to amyloid-β(1–42) toxicity	66
Caspase-3	Perinatally lethal (depend on genetic background)	Lack of apoptosis in neuroepithelial progenitor cells during development	22
Caspase-9	Embryonic lethal (a small percentage of caspase-9-knockout can survive to adult)	Lack of apoptosis in neuroepithelial progenitor cells during development	23,99
Caspase-11	Normal	Resistant to apoptosis induced by ischaemic brain injury	20
Caspase-12	Normal	Cortical neurons are resistant to amyloid-β toxicity	21

detected early after the onset of an ischaemic insult, in the penumbra where the insult is less severe and during reperfusion^{50,51}. It is possible that only neurons that maintain a minimum level of metabolic activity can undergo apoptosis, which is consistent with it being a cellular suicide programme.

Mitochondria might be important in transmitting apoptotic signals during ischaemia to induce caspase activation. There is strong evidence of caspase-3 activity in ischaemic brain⁵², which might be mediated by caspase-11 — a caspase that is specifically induced by ischaemic injury²⁰. Moreover, caspase inhibitors significantly attenuate ischaemic neuronal injury.

Although there is strong evidence for apoptosis in ischaemic brain injury, not all cells die by apoptosis. Among cells with typical apoptotic features, there are clearly cells with a swollen morphology and highly vacuolated features⁵³; thus, the death of a significant number of neurons in ischaemic brain is likely to occur through a non-caspase-mediated mechanism.

Neuronal cell death in Alzheimer's disease

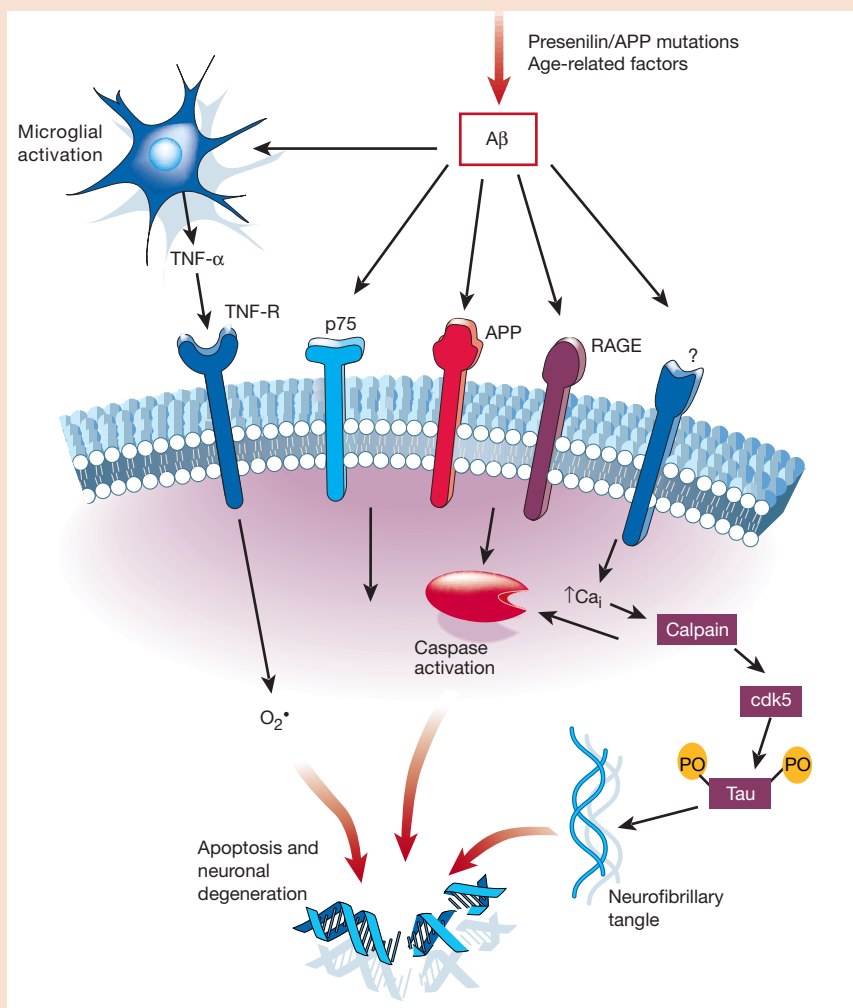
The relative contribution of apoptosis to neuronal loss in Alzheimer's disease is difficult to assess because of the chronic nature of the disease process, so that at any one time only a limited number of apoptotic neurons can be detected. Some neurons exhibit morphological features of apoptosis, but many degenerating neurons do not show evidence of apoptosis, suggesting that apoptosis might not be the only mechanism of degeneration in Alzheimer's disease^{54,55}.

The proximal cause of neurodegeneration in Alzheimer's disease is an actively debated issue that has become focused on several proteins implicated by genetics (Box 1). A central role for amyloid- β protein is supported by the effects of genetic mutations that cause

familial Alzheimer's disease⁵⁶, all of which predispose to amyloid deposition, and by the observation that amyloid- β can be neurotoxic *in vitro* and *in vivo*^{57,58}. The toxicity of abnormal structural forms of amyloid- β provides a unifying theme with other age-related neurodegenerative disorders characterized by the appearance of pathological protein structures, such as Parkinson's disease, Huntington's disease, frontotemporal dementia and amyotrophic lateral sclerosis (Fig. 3). The mechanism of amyloid- β neurotoxicity and its precise cellular locus of action are unsettled, but it has been shown that amyloid- β can induce oxidative stress and elevate intracellular Ca^{2+} concentration^{59,60} (Fig. 4). Amyloid- β might induce apoptosis⁶¹ by interacting with neuronal receptors, including the receptor for advanced glycation endproducts (RAGE), which can mediate free-radical production⁶², the p75 neurotrophin receptor, which can induce neuronal cell death⁶³, and the amyloid precursor protein, which can also induce neuronal cell death⁶⁴. These various amyloid- β -receptor interactions might activate several different cell-death-signalling pathways (Fig. 4). For example, amyloid- β can activate a set of immediate early genes similar to those induced by trophic factor withdrawal⁶⁵, and can activate caspases. Furthermore, neurons deficient in caspase-2 and caspase-12 have decreased vulnerability to amyloid- β toxicity^{21,66}, suggesting that selective caspase inhibition might be a potential therapeutic approach in Alzheimer's disease.

The identification of mutations in the presenilin genes as a major cause of early-onset familial Alzheimer's disease has provided a new approach to understanding the mechanism of neuronal cell death in Alzheimer's disease^{67–69}. Presenilin mutations increase the production of a 42-residue form of amyloid- β , the major constituent of plaques in the Alzheimer's disease brain⁷⁰. Several recent studies

Figure 4 Cellular pathways of amyloid- β protein neurotoxicity in Alzheimer's disease. Aggregated forms of amyloid- β interact with several different neuronal cell-surface receptors and with microglia, triggering signal transduction cascades that result in caspase activation, free-radical generation and Ca^{2+} influx. An increased intracellular concentration of Ca^{2+} (Ca_i) might activate calpain proteases which can, in turn, activate caspases and the tau protein kinase Cdk5.



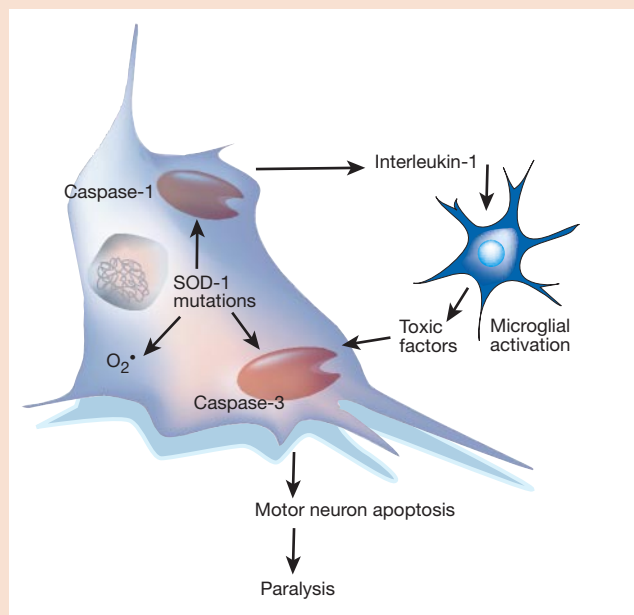


Figure 5 SOD-1 mutations activate cell death pathways in familial amyotrophic lateral sclerosis. SOD-1 mutations can activate caspase-1 and caspase-3, and might increase free-radical generation, leading to motor neuron apoptosis. The activation of caspase-1 leads to interleukin-1 production, which can induce a local microglial inflammatory response and increase the number of neurons affected.

suggest that presenilins might be λ -secretases, proteases that participate in the generation of amyloid- β , although this remains to be established definitively^{71,72}. Presenilin mutations can also increase neuronal vulnerability to apoptosis⁷³. But it remains to be determined whether these mechanisms contribute to neuronal cell death in Alzheimer's disease.

Activation of microglial cells is a prominent feature of the inflammatory response in the brain in Alzheimer's disease that is likely to contribute to neuronal cell death. Microglial activation is associated with amyloid plaques and can be induced experimentally by amyloid- β ⁷⁴. Amyloid- β -induced microglial activation results in the

secretion of TNF- α and other toxic factors that can induce neuronal apoptosis⁷⁵. Similar microglial-based mechanisms have been implicated in other neurodegenerative disorders. For example, a fibril-forming peptide derived from the prion protein induces neuronal apoptosis through microglial activation and the generation of reactive oxygen species⁷⁶. Microglial activation also has a central role in neuronal cell death associated with viral infections of the central nervous system. Microglia and macrophages are the predominant cell types infected by HIV in the brain⁷⁷, and induce the apoptosis of neurons and astrocytes in AIDS⁷⁸. Thus, pathological neuronal cell death might be a direct consequence of toxic insults such as amyloid- β , or an indirect consequence of a complex interaction between neurons, microglia and toxic factors.

Death from expanded polyglutamine repeats

The adult-onset neurodegenerative diseases caused by proteins with expanded polyglutamine tracts are characterized by a selective loss of specific neuronal subpopulations. Proteins with polyglutamine repeats can aggregate *in vitro* and form amyloid-like fibrils similar to the amyloid- β fibrils in Alzheimer's disease⁷⁹. Such aggregates are also observed in the brains of patients with Huntington's disease, spinocerebellar ataxia types 1 and 3, and dentatorubral-pallidoluysian atrophy⁸⁰. Ubiquitinated derivatives of the mutant proteins can be found in large intranuclear inclusions, and in Huntington's disease the number of inclusions is correlated with the length of the polyglutamine tract⁸¹. Ineffective clearance of polyglutamine expansion proteins by the ubiquitin-proteasome pathway might contribute to the formation of intranuclear inclusions⁸². Transgenic mice that express expanded polyglutamine-containing mutants of huntingtin and ataxin 1 also develop inclusions that appear at about the same time as the neurological deficits⁸³. Despite the correlation of the appearance of neuronal inclusions with disease in patients and transgenic mouse models, several studies have questioned their pathological significance^{84,85}. Moreover, the inhibition of inclusion formation can increase neuronal apoptosis *in vitro*, indicating that the formation of inclusions might be neuroprotective.

But there is evidence that expression of expanded polyglutamine tracts can result in the formation of small aggregates that do induce apoptosis⁸⁶. Apoptosis is mediated by the recruitment of the adaptor protein FADD (for Fas-associated death domain protein) and caspase-8, resulting in the activation of a caspase cascade. Furthermore, caspases might have a role in generating highly toxic fragments of

Box 1

Molecules implicated in the pathogenesis of Alzheimer's disease

Amyloid- β protein^{56,57}

1. This is the main component of senile plaques and cerebrovascular amyloid deposits.
2. All known genetic mutations that cause Alzheimer's disease predispose to amyloid deposition.
3. Individuals with trisomy 21, who carry an additional copy of the amyloid precursor protein gene, develop early-onset Alzheimer's disease.
4. Amyloid- β can be neurotoxic.

Tau¹⁰⁰

1. Hyperphosphorylated tau is the main component of neurofibrillary tangles.
2. Mutations in tau cause frontotemporal dementia with Parkinsonism associated with chromosome 17 (FTDP-17), suggesting that aberrant forms of tau can give rise to neurodegeneration. However, tau mutations have not been found in cases of Alzheimer's disease.
3. The number and distribution of neurofibrillary tangles are correlated with the degree of dementia in Alzheimer's disease.
4. Activation of protein kinase cdk5 might contribute to both tau phosphorylation and neuronal apoptosis¹⁰¹.

Presenilins⁶⁷⁻⁶⁹

1. Mutations in presenilin 1 and 2 are a major cause of early-onset familial Alzheimer's disease.
2. Presenilin mutations increase production of the 42-residue form of amyloid- β , which has a high propensity for forming amyloid fibrils.
3. Presenilins are required for amyloid- β production and might be γ -secretases.
4. Presenilin mutations increase neuronal vulnerability to apoptosis.

Apolipoprotein E⁵⁷

1. Inheritance of the $\epsilon 4$ allele is the most common known genetic risk factor for Alzheimer's disease after the age of 60.
2. The $\epsilon 4$ allele promotes the polymerization of amyloid- β into plaque-forming fibrils.
3. The $\epsilon 4$ allele might impair neuronal regeneration or promote oxidative stress.

proteins with expanded polyglutamine tracts. Support for this idea comes from studies on a transgenic model of Huntington's disease. A dominant-negative mutant of caspase-1, or the intracerebroventricular administration of a broad-spectrum caspase inhibitor, delays the onset and progression of pathology and prevents the appearance of a huntingtin cleavage product⁸⁷.

A central issue is the relative contribution of neuronal apoptosis to neurological deficits in polyglutamine expansion diseases and other age-related neurodegenerative disorders. Early-stage Huntington's disease patients develop characteristic motor deficits without evidence of striatal atrophy; striatal atrophy becomes prominent in later stages of the disease⁸⁸. Similarly, a transgenic mouse model of Huntington's disease exhibits motor symptoms in the absence of striatal atrophy or neuronal apoptosis⁸⁹. Furthermore, in a conditional huntingtin transgenic mouse, neuronal intranuclear inclusions and neurological deficits could be reversed by turning off expression of the mutant transgene⁹⁰. Thus, neuronal dysfunction, rather than cell death, might be responsible for early neurological deficits.

Mutations in superoxide dismutase and amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a progressive motor disease characterized by the degeneration of motor neurons in the spinal cord and brain, leading to paralysis. A major leap in understanding the disease mechanism came from the identification of mutations in the gene encoding superoxide dismutase (SOD-1) in familial ALS⁹¹. Transgenic mice that express mutant forms of SOD-1 show progressive motor neuron degeneration that is similar in many respects to that in the human disease⁹². In contrast, mice deficient in or overexpressing wild-type SOD-1 do not develop motor neuron disease. These findings suggest that mutant SOD-1 is somehow toxic to neurons. Although the mechanism of toxicity is not yet clearly established, it has been shown that mutant SOD-1 can form intraneuronal aggregates and induce oxidative stress, which is reminiscent of pathogenic mechanisms in Alzheimer's disease and polyglutamine repeat diseases⁹² (Fig. 5).

A role for apoptosis in familial ALS is suggested by the proapoptotic activity of mutant SOD-1 in cultured neural cell lines^{93,94}, and the neuroprotective effect of overexpressing Bcl-2 in mutant SOD-1-transgenic mice⁹⁵. Moreover, activated caspase-1 and caspase-3 can be detected in spinal cords of ALS patients and mutant SOD-1-transgenic mice^{94,96}. Importantly, the inhibition of caspase-1 activity delays disease progression in SOD-1-transgenic mice^{96,97}. Caspase-1 might predispose to neuronal cell death in two ways: by increasing production of the pro-inflammatory cytokine interleukin-1 and by directly activating caspase-3 (Fig. 5). All the evidence suggests that caspase activation may be an essential component of the pathology of ALS, and offers the possibility that early treatment with inhibitors targeted to specific caspases might arrest motor neuron apoptosis.

Conclusion

During the past decade there have been major advances in our understanding of the fundamental mechanisms of neuronal cell death. We now know that the key components of the apoptosis programme in neurons, like that of other cell types, are Apaf-1 and proteins in the Bcl-2 and caspase families. The regulation of apoptosis through interactions of Bcl-2 family members and caspase cascades has a major role in sculpting the developing brain. We are now beginning to understand how neurotrophins suppress apoptosis by regulating critical protein kinase cascades, such as the PI(3)K-Akt and MAP kinase pathways. Furthermore, not only are caspases important in regulating neuronal cell death during development, they might also mediate cell death in human neurodegenerative diseases. These exciting developments suggest that the targeted inhibition of apoptosis might be effective in the treatment of various neurodegenerative diseases.

Naturally, many fundamental questions remain to be answered. We do not yet understand exactly how a toxic stimulus, be it trophic

factor deprivation, ischaemic injury, amyloid- β peptide, mutant huntingtin or mutant SOD, triggers the activation of the apoptosis programme in neurons. This signal transduction process might be brief, as in trophic factor deprivation and ischaemic injury, or prolonged, as in neurons that express disease-causing mutant proteins. It will be important to define the molecular point of no return, when neurons become irreversibly committed to die. Obviously neuronal dysfunction might be initiated before neuronal degeneration, and from a therapeutic point of view a central question is whether the inhibition of neuronal cell death will result in healthy, normally functioning neurons. The answers to these questions and the design of rational therapeutic approaches will require a detailed understanding of how neurons survive and die in the brain. □

- Hamburger, V. & Levi-Montalcini, R. *J. Exp. Zool.* **111**, 457–502 (1949).
- Purves, D. *Body and Brain: A Trophic Theory of Neural Connections* (Harvard Press, Cambridge, Massachusetts, 1988).
- Martin, D. P. *et al.* Inhibitors of protein synthesis and RNA synthesis prevent neuronal death caused by nerve growth factor deprivation. *J. Cell Biol.* **106**, 829–844 (1988).
- Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. Genetics of programmed cell death in *C. elegans*: past, present and future. *Trends Genet.* **14**, 410–416 (1998).
- Gagliardini, V. *et al.* Prevention of vertebrate neuronal death by the *crmA* gene. *Science* **263**, 826–828 (1994).
- Merry, D. E. & Korsmeyer, S. J. Bcl-2 gene family in the nervous system. *Annu. Rev. Neurosci.* **20**, 245–267 (1997).
- Garcia, I., Martinou, J. C., Tsujimoto, Y. & Martinou, J. C. Prevention of programmed cell death of sympathetic neurons by the *bcl-2* proto-oncogene. *Science* **258**, 302–304 (1992).
- Martinou, J. C. *et al.* Overexpression of Bcl-2 in transgenic mice protects neurons from naturally occurring cell death and experimental ischemia. *Neuron* **13**, 1017–1030 (1994).
- Dubois-Dauphin, M., Frankowski, H., Tsujimoto, Y., Huarte, J. & Martinou, J. C. Neonatal motoneurons overexpressing the *bcl-2* protooncogene in transgenic mice are protected from axotomy-induced cell death. *Proc. Natl Acad. Sci. USA* **91**, 3309–3313 (1994).
- Sagot, Y. *et al.* Bcl-2 overexpression prevents motoneuron cell body loss but not axonal degeneration in a mouse model of a neurodegenerative disease. *J. Neurosci.* **15**, 7727–7733 (1995).
- Weis, D. J., Sorenson, C. M., Shutter, J. R. & Korsmeyer, S. J. Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair. *Cell* **75**, 229–240 (1993).
- Michaelidis, T. M. *et al.* Inactivation of *bcl-2* results in progressive degeneration of motoneurons, sympathetic and sensory neurons during early postnatal development. *Neuron* **17**, 75–89 (1996).
- Gonzalez-Garcia, M. *et al.* *bcl-x* is expressed in embryonic and postnatal neural tissues and functions to prevent neuronal cell death. *Proc. Natl Acad. Sci. USA* **92**, 4304–4308 (1995).
- Motoyama, N. *et al.* Massive cell death of immature hematopoietic cells and neurons in Bcl-x-deficient mice. *Science* **267**, 1506–1510 (1995).
- Deckwerth, T. L. *et al.* BAX is required for neuronal death after trophic factor deprivation and during development. *Neuron* **17**, 401–411 (1996).
- Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome *c*-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
- Cecconi, F., Alvarez-Bolado, G., Meyer, B. I., Roth, K. A. & Gruss, P. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
- Cryns, V. & Yuan, J. Proteases to die for. *Genes Dev.* **12**, 1551–1570 (1998).
- Li, P. *et al.* Cytochrome *c* and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
- Kang, S. J. *et al.* Dual role of caspase-11 in mediating activation of caspase-1 and caspase-3 under pathological conditions. *J. Cell Biol.* **149**, 613–622 (2000).
- Nakagawa, T. *et al.* Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid- β . *Nature* **403**, 98–103 (2000).
- Kuida, K. *et al.* Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* **384**, 368–372 (1996).
- Kuida, K. *et al.* Reduced apoptosis and cytochrome *c*-mediated caspase activation in mice lacking caspase 9. *Cell* **94**, 325–337 (1998).
- Barbacid, M. Structural and functional properties of the TRK family of neurotrophin receptors. *Ann. NY Acad. Sci.* **766**, 442–458 (1995).
- Fruman, D. A., Meyers, R. E. & Cantley, L. C. Phosphoinositide kinases. *Annu. Rev. Biochem.* **67**, 481–507 (1998).
- Yao, R. & Cooper, G. M. Regulation of the Ras signaling pathway by GTPase-activating protein in PC12 cells. *Oncogene* **11**, 1607–1614 (1995).
- Philpott, K. L., McCarthy, M. J., Klippel, A. & Rubin, L. L. Activated phosphatidylinositol 3-kinase and Akt kinase promote survival of superior cervical neurons. *J. Cell Biol.* **139**, 809–815 (1997).
- Datta, S. R., Brunet, A. & Greenberg, M. E. Cellular survival: a play in three Akts. *Genes Dev.* **13**, 2905–2927 (1999).
- Yang, E. *et al.* Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. *Cell* **80**, 285–291 (1995).
- Datta, S. R. *et al.* Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell* **91**, 231–241 (1997).
- Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
- Du, K. & Montminy, M. CREB is a regulatory target for the protein kinase Akt/PKB. *J. Biol. Chem.* **273**, 32377–32379 (1998).
- Riccio, A., Ahn, S., Davenport, C. M., Blendy, J. A. & Ginty, D. D. Mediation by a CREB family transcription factor of NGF-dependent survival of sympathetic neurons. *Science* **286**, 2358–2361 (1999).
- Kane, L. P., Shapiro, V. S., Stokoe, D. & Weiss, A. Induction of NF- κ B by the Akt/PKB kinase. *Curr. Biol.* **9**, 601–604 (1999).

35. Bonni, A. *et al.* Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science* **286**, 1358–1362 (1999).
36. Koike, T., Martin, D. P. & Johnson, E. M. Jr Role of Ca^{2+} channels in the ability of membrane depolarization to prevent neuronal death induced by trophic-factor deprivation: evidence that levels of internal Ca^{2+} determine nerve growth factor dependence of sympathetic ganglion cells. *Proc. Natl Acad. Sci. USA* **86**, 6421–6425 (1989).
37. Mao, Z., Bonni, A., Xia, F., Nadal-Vicens, M. & Greenberg, M. E. Neuronal activity-dependent cell survival mediated by transcription factor MEF2. *Science* **286**, 785–790 (1999).
38. Vaillant, A. R. *et al.* Depolarization and neurotrophins converge on the phosphatidylinositol 3-kinase-Akt pathway to synergistically regulate neuronal survival. *J. Cell Biol.* **146**, 953–966 (1999).
39. Xia, Z., Dickens, M., Raingeaud, J., Davis, R. J. & Greenberg, M. E. Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science* **270**, 1326–1331 (1995).
40. Estus, S. *et al.* Altered gene expression in neurons during programmed cell death: identification of *c-jun* as necessary for neuronal apoptosis. *J. Cell Biol.* **127**, 1717–1727 (1994).
41. Ham, J. *et al.* A *c-Jun* dominant negative mutant protects sympathetic neurons against programmed cell death. *Neuron* **14**, 927–939 (1995).
42. Imaizumi, K. *et al.* The cell death-promoting gene DP5, which interacts with the BCL2 family, is induced during neuronal apoptosis following exposure to amyloid beta protein. *J. Biol. Chem.* **274**, 7975–7981 (1999).
43. Putcha, G. V., Deshmukh, M. & Johnson, E. M. Jr BAX translocation is a critical event in neuronal apoptosis: regulation by neuroprotectants, BCL-2, and caspases. *J. Neurosci.* **19**, 7476–7485 (1999).
44. Deshmukh, M. & Johnson, E. M. Jr Evidence of a novel event during neuronal death: development of competence-to-die in response to cytoplasmic cytochrome *c*. *Neuron* **21**, 695–705 (1998).
45. Dechant, G. & Barde, Y. A. Signalling through the neurotrophin receptor p75NTR. *Curr. Opin. Neurobiol.* **7**, 413–418 (1997).
46. Frade, J. M., Rodriguez-Tebar, A. & Barde, Y. A. Induction of cell death by endogenous nerve growth factor through its p75 receptor. *Nature* **383**, 166–168 (1996).
47. Casaccia-Bonnel, P., Carter, B. D., Dobrowsky, R. T. & Chao, M. V. Death of oligodendrocytes mediated by the interaction of nerve growth factor with its receptor p75. *Nature* **383**, 716–719 (1996).
48. Bamji, S. X. *et al.* The p75 neurotrophin receptor mediates neuronal apoptosis and is essential for naturally occurring sympathetic neuron death. *J. Cell Biol.* **140**, 911–923 (1998).
49. Davey, F. & Davies, A. M. TrkB signalling inhibits p75-mediated apoptosis induced by nerve growth factor in embryonic proprioceptive neurons. *Curr. Biol.* **8**, 915–918 (1998).
50. Li, Y., Chopp, M., Jiang, N., Zhang, Z. G. & Zaloga, C. Induction of DNA fragmentation after 10 to 120 minutes of focal cerebral ischemia in rats. *Stroke* **26**, 1252–1257; discussion 1257–1258 (1995).
51. Charriaut-Marlangue, C. *et al.* Apoptosis and necrosis after reversible focal ischemia: an in situ DNA fragmentation analysis. *J. Cereb. Blood Flow Metab.* **16**, 186–194 (1996).
52. Namura, S. *et al.* Activation and cleavage of caspase-3 in apoptosis induced by experimental cerebral ischemia. *J. Neurosci.* **18**, 3659–3668 (1998).
53. Martin, L. J. *et al.* Neurodegeneration in excitotoxicity, global cerebral ischemia, and target deprivation: a perspective on the contributions of apoptosis and necrosis. *Brain Res. Bull.* **46**, 281–309 (1998).
54. Su, J. H., Anderson, A. J., Cummings, B. J. & Cotman, C. W. Immunohistochemical evidence for apoptosis in Alzheimer's disease. *Neuroreport* **5**, 2529–2533 (1994).
55. Troncoso, J. C., Sukhov, R. R., Kwas, C. H. & Koliatsos, V. E. In situ labeling of dying cortical neurons in normal aging and in Alzheimer's disease: correlations with senile plaques and disease progression. *J. Neuropathol. Exp. Med.* **55**, 1134–1142 (1996).
56. Selkoe, D. J. Alzheimer's disease: genotypes, phenotypes and treatments. *Science* **275**, 630–631 (1997).
57. Yankner, B. A. Mechanisms of neuronal degeneration in Alzheimer's disease. *Neuron* **16**, 921–932 (1996).
58. Geula, G. *et al.* Aging renders the brain vulnerable to amyloid β -protein neurotoxicity. *Nature Med.* **4**, 827–831 (1998).
59. Behl, C., Davis, J. B., Lesley, R. & Schubert, D. Hydrogen peroxide mediates amyloid β protein toxicity. *Cell* **77**, 817–827 (1994).
60. Mattson, M. P., Tomaselli, K. J. & Rydel, R. E. Calcium-depleting and neurodegenerative effect of aggregate beta-amyloid peptide are attenuated by basic GGF. *Brain Res.* **621**, 35–49 (1993).
61. Loo, D. T. *et al.* Apoptosis is induced by β -amyloid in cultured central nervous system neurons. *Proc. Natl Acad. Sci. USA* **90**, 7951–7955 (1993).
62. Yan, S. D. *et al.* RAGE and amyloid- β peptide neurotoxicity in Alzheimer's disease. *Nature* **382**, 685–691 (1996).
63. Yaar, M. *et al.* Binding of beta-amyloid to the p75 neurotrophin receptor induces apoptosis. A possible mechanism for Alzheimer's disease. *J. Clin. Invest.* **100**, 2333–2340 (1997).
64. Lorenzo, A. *et al.* Amyloid- β interacts with the amyloid precursor protein: a potential toxic mechanism in Alzheimer's disease. *Nature Neurosci.* **3**, 460–464 (2000).
65. Estus, S. *et al.* Aggregated amyloid-beta protein induces cortical neuronal apoptosis and concomitant 'apoptotic' pattern of gene induction. *J. Neurosci.* **17**, 7736–7745 (1997).
66. Troy, C. M. *et al.* Caspase-2 mediates neuronal cell death induced by beta-amyloid. *J. Neurosci.* **20**, 1386–1392 (2000).
67. Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760 (1995).
68. Levy-Lahad, E. *et al.* Candidate gene for chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977 (1995).
69. Price, D. L., Tanzi, R. E., Borchelt, D. R. & Sisodia, S. S. Alzheimer's disease: genetic studies and transgenic models. *Annu. Rev. Genet.* **32**, 461–493 (1998).
70. Scheuner, D. *et al.* Secreted amyloid β -protein similar to that in the senile plaques of Alzheimer's disease is increased in *vivo* by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Med.* **2**, 864–870 (1996).
71. Wolfe, M. S. *et al.* Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ -secretase activity. *Nature* **398**, 513–517 (1999).
72. Li, Y. M. *et al.* Photoactivated γ -secretase inhibitors directed to the active site covalently label presenilin 1. *Nature* **405**, 689–694 (2000).
73. Mattson, M. P., Guo, Q., Furukawa, K. & Pedersen, W. A. Presenilins, the endoplasmic reticulum, and neuronal apoptosis in Alzheimer's disease. *J. Neurochem.* **70**, 1–14 (1998).
74. Giulian, D. *et al.* Specific domains of beta-amyloid from Alzheimer plaque elicit neuron killing in human microglia. *J. Neurosci.* **16**, 6021–6037 (1996).
75. Tan, J. *et al.* Microglial activation resulting from CD40-CD40L interaction after beta-amyloid stimulation. *Science* **286**, 2352–2355 (1999).
76. Brown, D. R., Schmidt, B. & Kretschmer, H. A role of microglia and host prion protein in neurotoxicity of prion protein fragment. *Nature* **380**, 345–347 (1996).
77. Gonzalez-Scarano, F. B. Microglia as mediators of inflammatory and degenerative diseases. *Annu. Rev. Neurosci.* **22**, 219–240 (1999).
78. Ohagen, A. *et al.* Apoptosis induced by infection of primary brain cultures with diverse human immunodeficiency virus type 1 isolates: evidence for a role of the envelope. *J. Virol.* **73**, 897–906 (1999).
79. Scherzinger, E. *et al.* Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **90**, 549–558 (1997).
80. Lunkes, A. M. Polyglutamines, nuclear inclusions and neurodegeneration. *Nature Med.* **3**, 1201–1202 (1997).
81. DiFiglia, M. *et al.* Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990–1993 (1997).
82. Orr, H. T. & Zoghbi, H. Y. Reversing neurodegeneration: a promise unfolds. *Cell* **101**, (2000).
83. Davies, S. W. *et al.* Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537–548 (1997).
84. Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. E. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55–66 (1998).
85. Klement, I. A. *et al.* Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* **95**, 41–53 (1998).
86. Sanchez, L. *et al.* Caspase-8 is required for cell death induced by expanded polyglutamine repeats. *Neuron* **22**, 623–633 (1999).
87. Ona, V. O. *et al.* Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. *Nature* **399**, 204–205, 207 (1999).
88. Vonsattel, J. P. *et al.* Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.* **44**, 559–577 (1985).
89. Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493–506 (1996).
90. Yamamoto, A., Lucas, J. J. & Hen, R. Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. *Cell* **101**, 57–66 (2000).
91. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59–62 (1993).
92. Cleveland, D. W. From Charcot to SOD1: mechanisms of selective motor neuron death in ALS. *Neuron* **24**, 515–520 (1999).
93. Rabizadeh, S. *et al.* Mutations associated with amyotrophic lateral sclerosis convert superoxide dismutase from an antiapoptotic gene to a proapoptotic gene: studies in yeast and neural cells. *Proc. Natl Acad. Sci. USA* **92**, 3024–3028 (1995).
94. Pasinelli, P. *et al.* Caspase-1 is activated in neuronal cells and tissue with amyotrophic lateral sclerosis-associated mutations in copper-zinc superoxide dismutase. *Proc. Natl Acad. Sci. USA* **95**, 15763–15768 (1998).
95. Kostic, V., Jackson-Lewis, V., de Bilbao, F., Dubois-Dauphin, M. & Przedborski, S. Bcl-2: prolonging life in a transgenic mouse model of familial amyotrophic lateral sclerosis. *Science* **277**, 559–562 (1997).
96. Li, M., Ona, V. O., Guegan, C. & Chen, M. Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model. *Science* **288**, 335–339 (2000).
97. Friedlander, R. M., Brown, R. H., Gagliardini, V., Wang, J. & Juan, J. Inhibition of ICE slows ALS in mice. *Nature* **388**, 31 (1997).
98. Liu, X. H. *et al.* Mice deficient in interleukin-1 converting enzyme are resistant to neonatal hypoxic-ischemic brain damage. *J. Cereb. Blood Flow Metab.* **19**, 1099–1108 (1999).
99. Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell* **94**, 339–352 (1998).
100. Lee, V. M. Y. & Trojanowski, J. Q. Neurodegenerative tauopathies: human disease and transgenic mouse models. *Neuron* **24**, 507–510 (1999).
101. Patrick, G. N. *et al.* Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration. *Nature* **402**, 615–622 (1999).

Acknowledgements

We thank M. Greenberg, M. Moskowitz, M. Deshmukh and E. Johnson for critical readings of the manuscript. This work was supported by grants from the NIH (to B.Y. and J.Y.), a grant from the American Heart Association (to J.Y.), a Zenith Award from the Alzheimer's Association (to B.Y.) and a NIH MRRC core grant.

From bench to clinic with apoptosis-based therapeutic agents

Donald W. Nicholson

Merck Frosst Centre for Therapeutic Research, Merck Research Laboratories, PO 1005 Pointe Claire-Dorval, Quebec, Canada H9R 4P8
(e-mail: donald_nicholson@merck.com)

A retrospective look at the basis of human disease pathogenesis almost always reveals an apoptotic component that either contributes to disease progression or accounts for it. What makes this field particularly exciting is the breadth of therapeutic opportunities that are on offer. The pace of apoptosis research has raised expectations that therapeutics will follow soon. But many of the organizations that are best placed to take advantage of these discoveries consider the ability to modulate the life or death of a cell for the purpose of disease treatment as perhaps being 'too good to be true'. Nevertheless, practical therapeutics that modulate apoptosis will no doubt appear in the clinic or on the shelf in the next few years.

The human body is composed of approximately 10^{14} cells, each of which is capable of committing suicide by apoptosis. Not surprisingly (although just recently understood), this process has inherent weaknesses that, when compromised, can result in inappropriate cell death (either too little or too much) and disease pathogenesis. Other articles in this issue have highlighted the complex role that apoptosis has in the homeostasis of multicellular organisms; here I describe the immense potential for therapeutic use that exists in modulating apoptosis for the treatment of human diseases.

Technological power meets apoptosis

The springboard for many of these discoveries has been the genetic blueprint of the cell death pathway, which was originally defined in the nematode *Caenorhabditis elegans*¹. Each of the central components of this pathway were found to have mammalian counterparts that often existed as multigene families. Their rapid discovery was empowered by the birth of computational biology (formerly termed 'bioinformatics'), which evolved alongside the advances made in genomics resources. For example, public-domain expressed sequence tag (EST) collections were used to identify key members of the Bcl-2, caspase, tumour necrosis factor (TNF) receptor and adapter molecule families — often simultaneously in multiple laboratories — and the now-completed human genome sequence will no doubt fill in the gaps. Understanding the function and relationship between these molecules was aided by an ever-evolving set of impressive cell biology tools, including interaction analysis with yeast two-hybrid systems, and knockout mice whose phenotype frequently clarified an otherwise complex biology. In addition, nearly 20 three-dimensional structures for biomolecules involved in apoptosis have been resolved using X-ray or nuclear magnetic resonance techniques, yielding a better understanding of their *modus operandi* and their functional interrelationships, as well as the design of specific inhibitors. The therapeutic landscape has also witnessed major changes. Antisense-based therapies have been made more viable by better oligonucleotide chemistry, resulting in increased metabolic stability, cell penetration and fewer side effects. Injectable recombinant biologicals are demonstrating clinical efficacy and are becoming more broadly accepted as legitimate therapeutic agents. Last but not least,

combinatorial chemistry and rapid analogue synthesis — techniques by which large numbers of chemical permutations can be quickly assembled — have greatly accelerated the pace at which potential drug candidates can be generated. These remarkable accomplishments have positioned the apoptosis field for what is hoped will be a smooth transition from bench to clinic.

Opportunities and limitations

Even a quick glance at the molecular components of the cell death pathway (see the other articles in this issue) reveals many opportunities for apoptosis modulation. With this plethora of opportunities unfortunately comes a number of practical limitations, not the least of which is the practical issue that the cell death pathway as currently known contains very few conventional drug targets, such as enzymes and small-ligand receptors. Attention has therefore focused on other strategies to affect the proteinaceous components of the apoptotic pathway.

Modulating the expression of key molecular components of the cell death machinery is an attractive and obvious strategy. But whereas gene and antisense therapy seem the most viable approaches at present to alter gene expression, and antisense Bcl-2 shows promise in the treatment of cancer, these technologies are still in their infancy.

An alternative is to interfere with specific protein–protein interactions inside or outside the cell. But despite intensive research, small molecules that interfere with specific protein–protein interactions are almost unheard of because the surface areas between interacting polypeptides are large and difficult to disrupt. The 'critical interaction' strategy, where small-molecule inhibitors disrupt only key sites of binding between interacting polypeptides, is sound in principle but has been lacking in practice. Even so, small molecules with a high affinity for the binding cleft of the Bcl-2 homology (BH) domain BH3 on the surface of Bcl-2 seem to induce apoptosis in cultured cells².

For recombinant protein strategies, a different set of equally complex issues arise. For example, there is always the possibility that the cells of the immune system will develop autoantibodies against recombinant proteins and, furthermore, their use is mostly limited to extracellular targets because large proteins do not readily enter the cell.

Opportunities are scarce for the development of

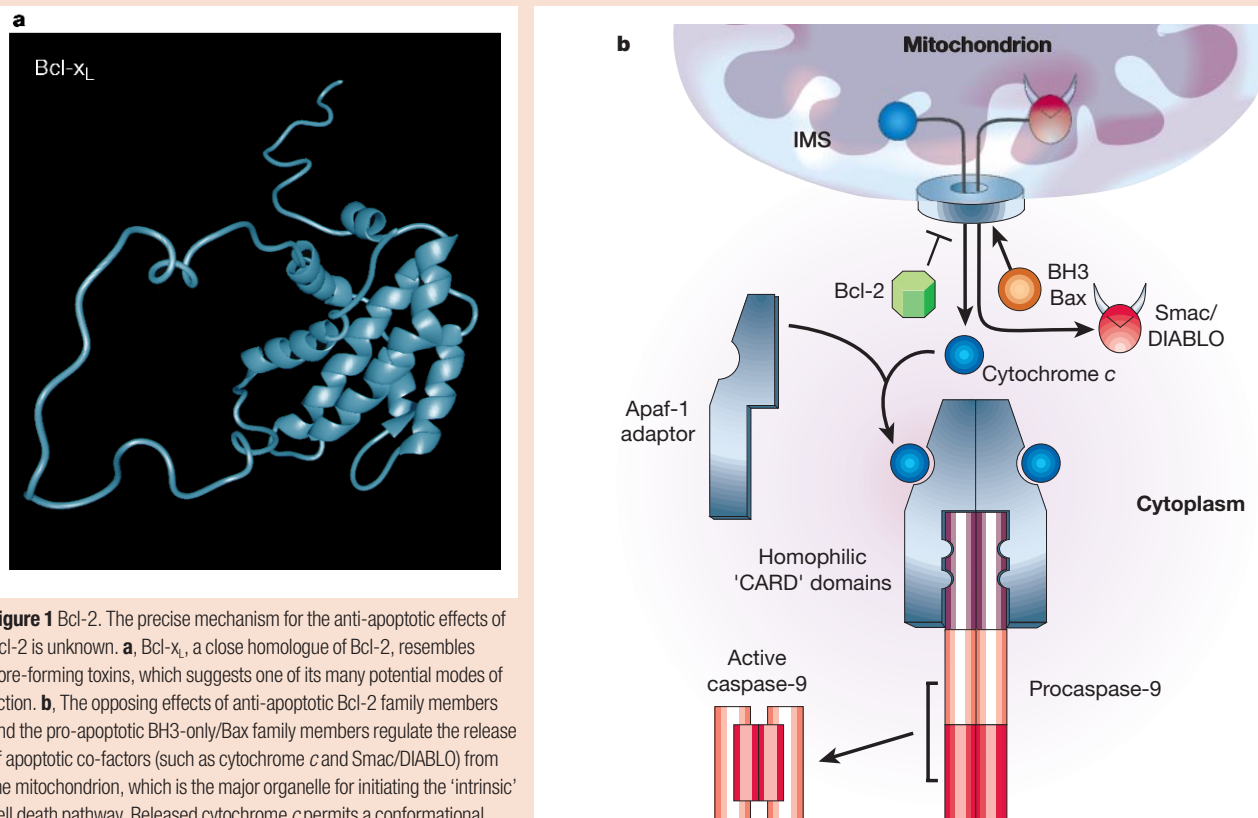


Figure 1 Bcl-2. The precise mechanism for the anti-apoptotic effects of Bcl-2 is unknown. **a**, Bcl-x_L, a close homologue of Bcl-2, resembles pore-forming toxins, which suggests one of its many potential modes of action. **b**, The opposing effects of anti-apoptotic Bcl-2 family members and the pro-apoptotic BH3-only/Bax family members regulate the release of apoptotic co-factors (such as cytochrome *c* and Smac/DIABLO) from the mitochondrion, which is the major organelle for initiating the 'intrinsic' cell death pathway. Released cytochrome *c* permits a conformational change in the cytosolic adaptor molecule, Apaf-1, which allows it to recruit and oligomerize caspase-9 via homophilic caspase recruitment domains (CARDs). Caspase-9 becomes activated and launches the apoptotic pathway. Some evidence indicates that Bcl-2 might also sequester Apaf-1 and that it is released or displaced by the pro-apoptotic (BH3-only/Bax) family members. IMS, mitochondrial intermembrane space.

pharmaceutical therapeutics to modulate the apoptotic pathway; the exception being organic small-molecule inhibitors of caspase activity. But there is no precedent of human therapeutics that successfully target cysteine proteases owing to the difficulties in developing electrophiles that are specific enough that for them not to attack other biological nucleophiles.

On top of all these practical limitations come the theoretical ones. For example, can apoptosis be selectively modulated in one organ or cell type without adverse effects on other key systems? The redundancy that exists in several of the multigene families suggests that this might just be possible, but it has yet to be proven *in vivo*. Similarly, if cells are salvaged by inhibiting apoptosis, will they be functional? In all likelihood this will depend on the cell type, its context and the degree of cellular injury inflicted on it. Nevertheless, impressive advances have been made to modulate apoptosis *in vivo* and although modulating apoptosis will not be a panacea for all of humanity's ailments, the wealth of opportunities merit a persistent effort.

Apoptotic-modulating therapies

Apoptosis-modulating therapeutics are now finally in human clinical trials or are on the brink, having shown efficacy in preclinical animal models. This is significant progress, given that the field as a whole is only about a decade old. The following examples are three of the most advanced and promising opportunities which, it is hoped, will set the stage and establish precedents for many future therapeutics. Each targets a different biochemical component of the cell death pathway and, surprisingly, each represents a different therapeutic modality (for example, disruption of gene function with antisense oligonucleotides (Bcl-2), recombinant biologicals (TRAIL, for tumour necrosis factor-related apoptosis-inducing ligand) and classical organic pharmaceuticals (caspases)). Furthermore, these entities tackle the complex issues of how, on the one hand, to activate

apoptosis selectively for the treatment of disorders where there is insufficient apoptosis, such as cancer (Bcl-2 and TRAIL), and how, on the other hand, to block cell death selectively for the treatment of diseases in which excessive apoptosis occurs and needs to be attenuated, such as in neurodegeneration (caspases).

Bcl-2 antisense

The Bcl-2 family of proteins are among the most studied molecules in the apoptotic pathway^{3,4} (Fig. 1a). Although the precise mechanism by which they function as anti-apoptotic molecules remains unclear, the cellular stoichiometry of Bcl-2 family members compared with their pro-apoptotic ('Bax/BH3-only') homologues clearly defines the vulnerability of cells to most, but not all, death stimuli (Fig. 1b). Bcl-2 itself was first identified in B-cell lymphomas in which the genetic lesion was a translocation of the Bcl-2 gene to the control of the immunoglobulin promoter (t(14;18)). The resulting overexpression of Bcl-2 retards the normal course of apoptotic cell death that otherwise occurs to maintain B-cell homeostasis, resulting in B-cell accumulation and follicular lymphoma. This observation showed that cancers do not strictly arise from unrestrained cell proliferation, but could also be due to insufficient apoptotic turnover. It further suggested that a decrease in Bcl-2 levels or the inhibition of Bcl-2 activity might provoke apoptosis or at least sensitize cells to apoptotic death. This has been verified in a large number of studies *in vitro* and is consistent with the phenotype of both Bcl-2-null mice — which develop normally but eventually display marked lymphoid apoptosis, as well as melanocyte, neuronal and intestinal lesions and terminal kidney disease — and Bcl-2 transgenic animals, which accumulate mature B lymphocytes and recapitulate much of the human phenotype^{5,6}. In addition to follicular lymphomas, Bcl-2 levels are elevated in a broad range of other human cancers, indicating that this molecule might have a role in raising the apoptotic threshold in a

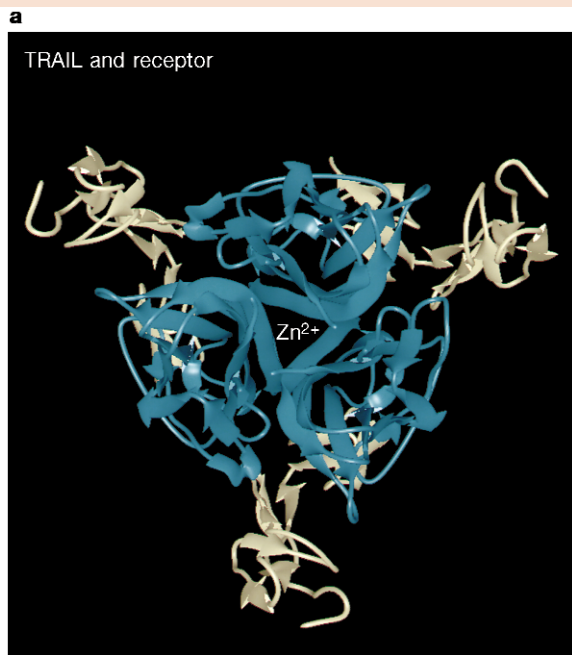
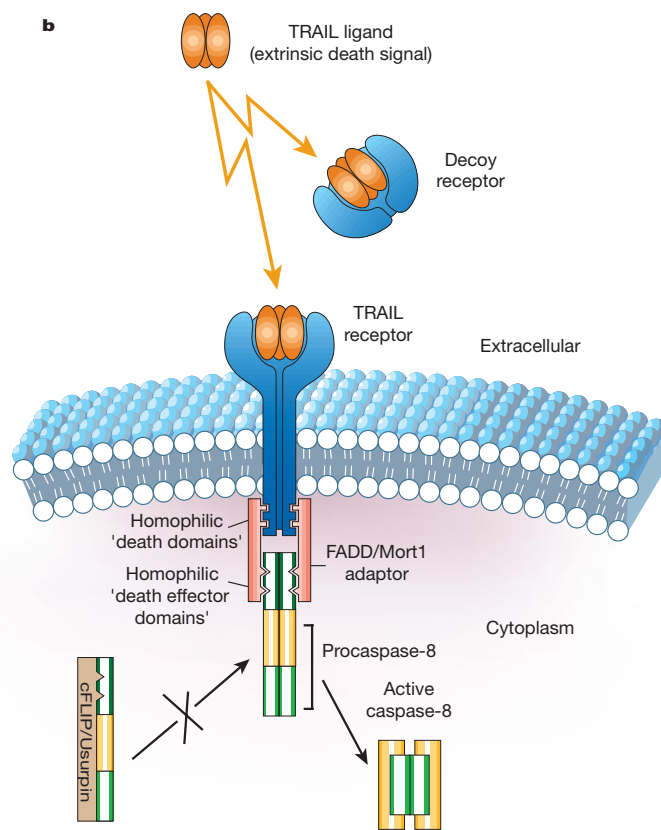


Figure 2 TRAIL. The key elements of the TRAIL signalling pathway are conserved in other members of the TNF 'death receptor' family, such as CD95 (Apo-1/Fas).

a. The trimerized TRAIL ligand (blue) is shown from a top-down view nestled into a trimer of the DR5/TRAIL-R2 receptor ectodomain (light brown). Coordinated Zn^{2+} seems to have a key role in conferring the appropriate signalling conformation.

b. The 'extrinsic' cell death pathway is launched by death-receptor ligands that trigger caspase-8 oligomerization and proximity-induced autoproteolytic activation via adaptor molecules such as FADD/Mort1. Activation is regulated by decoy receptors, which preclude the binding of TRAIL to the functional receptor, and dominant-negative pseudo-caspases such as c-FLIP/Usurpin, which prevent the recruitment of the caspase-8 proenzyme into the receptor complex.



broad spectrum of cancerous disorders. This would bode well for a Bcl-2-directed therapeutic, although it is likely — given the size of this multigene family — that other Bcl-2 family members will be the dominant anti-apoptotic effector in some types of cancer.

In the absence of a clearly defined biochemical mechanism of action for this family of cell-death regulatory proteins (for which conventional inhibitors could therefore be developed), gene therapy and antisense approaches have become the logical alternative. This has come the furthest for the 18-mer all-phosphorothioate Bcl-2 antisense oligonucleotide, G-3139 (Genta). This molecule (5'-d(P-thio)TCT-CCC-AGC-GTG-CGC-CAT-3') targets the first six codons of the human Bcl-2 open reading frame. The antisense binds to the Bcl-2 mRNA, thus precluding it from translation into Bcl-2 protein and targeting the message for degradation. Phosphorothioate bonds improve metabolic stability within the antisense oligonucleotide. This tips the balance between pro-apoptotic and anti-apoptotic family members in favour of pro-apoptotic members, resulting in apoptosis.

Preclinical data

In preclinical animal models, xenotransplantations of human tumours into mice with severe combined immunodeficiency (SCID) are markedly affected by continuous subcutaneous infusion of the Bcl-2 antisense G-3139 but not by the appropriate controls such as reverse, scrambled or mismatched oligonucleotides. For example, a marked decrease in tumour growth (more than 90%) of Merkel cell carcinomas (an aggressive neuroendocrine skin cancer with high metastatic potential) that were xenografted into SCID mice was observed after 28 days of treatment, and efficacy was superior to treatment with cisplatin⁷. Importantly, the apoptotic death of tumour cells in other models is enhanced by co-administration with standard chemotherapeutic agents, indicating, as might have been

predicted, that lower Bcl-2 levels decrease the apoptotic threshold of these cells and make them chemosensitive (or at least reverse Bcl-2-mediated chemoresistance). For example, human melanomas grown in SCID mice were ablated by a combination of the Bcl-2 antisense G-3139 and dacarbazine (the only single agent currently approved by the US Food and Drug Administration (FDA) for treating metastatic melanoma)⁸.

Clinical data

This has translated favourably into the clinic in phase I/IIa studies, which demonstrated that 43% of treated patients (6 of 14) showed clear anti-tumour responses associated with decreased Bcl-2 levels and increased apoptosis in melanoma biopsies⁹. G-3139 is currently in phase III clinical trials for malignant melanoma in the United States, for which it has been awarded fast-track priority by the FDA.

Positive results are also emerging for other human cancers, indicating that this approach might indeed have wide applicability. Encouraging results have been described for the treatment of non-Hodgkin's lymphoma. In a phase I study in relapsed patients with Bcl-2-positive lymphomas, disease stabilization was seen in 43% (9 of 21) and improvements were seen in 14% (3 of 21, including one complete responder)¹⁰. Phase II clinical studies are now underway to improve efficacy by combining the Bcl-2 antisense oligonucleotide with conventional chemotherapeutics, such as cyclophosphamide. Similarly, combination approaches using chemotherapeutics with antisense Bcl-2 are being examined in the clinic for relapsed small-cell lung carcinoma (paclitaxel), hormone-resistant metastatic prostate cancer (mitoxantrone), breast cancer (docetaxel), colorectal cancer (irinotecan) and relapsed acute leukaemia (fludarabine and cytosine arabinoside).

Bcl-2 is only one of the molecules within the apoptotic cell death pathway that could, in principle, be manipulated for therapeutic

gain. Others that might soon follow include Bcl-2-related proteins, such as Bcl-x_L. This protein can be modulated simultaneously with Bcl-2 by means of a bi-specific antisense oligonucleotide that targets a region of sequence identity between their respective messenger RNAs (which is not contained in the pro-apoptotic splice variant, Bcl-x_S)¹¹. In principle, this strategy could further accentuate the chemosensitivity of cancer cells by targeting two key anti-apoptotic proteins instead of just one. Similarly, Bcl-x antisense oligonucleotides that can shift pre-mRNA splicing away from the formation of the anti-apoptotic variant Bcl-x_L to the pro-apoptotic variant Bcl-x_S can sensitize cells to apoptotic stimuli^{12,13}.

FLIP and survivin antisense

An alternative opportunity for antisense-based therapy may be provided by c-FLIP, which is a naturally occurring dominant-negative antagonist of death-receptor signal transduction (see Fig. 2b). Inhibition of c-FLIP might be useful in the treatment of carcinomas that have acquired resistance to CD95 (Apo-1/Fas)-dependent killing, as has been demonstrated in tissue culture with highly malignant adenocarcinomas originating from cholangiocytes¹⁴.

Finally, one of the most prominent gene products associated with a wide variety of human cancers is survivin, a member of the inhibitor-of-apoptosis (IAP) family of proteins (see Fig. 3b). Antisense oligonucleotides directed against this molecule can induce spontaneous apoptosis in lung cancer cells, malignant melanomas and other cancer cell types. Overall, there are many opportunities to reset the apoptotic threshold of cancerous cells using antisense technology, particularly for acute treatment. In practice, their true utility will depend on many subtleties within the cell death pathway and how exactly it has been corrupted in specific disease pathologies.

Recombinant TRAIL (Apo-2L)

Cell death signals are commuted via two major biochemical routes in mammalian cells. The 'intrinsic' pathway, which responds to most pro-apoptotic signals, is modulated by the interplay between Bcl-2 and Bax/BH3-only family members and involves cues emanating largely via the mitochondrion (see Fig. 1b). The 'extrinsic' pathway, in contrast, is triggered by the ligation of 'death' receptors belonging to the TNF-receptor superfamily^{15,16} (Fig. 2a). Activation of these receptors results in recruitment of caspase zymogens into oligomeric complexes and triggers their proteolytic activation^{17,18} (Fig. 2b). 'Death receptors' all contain homophilic 'death domains' within their cytoplasmic extensions, which serve to recruit adapter proteins that in turn recruit procaspase-8. CD95, the prototypical death receptor, does so via FADD (for Fas-associated death domain protein), a bipartite adapter that directly bridges the CD95-ligand-ligated receptor with the caspase-8 proenzyme. TNF receptor 1 (which binds TNF- α) uses a similar mechanism except that an additional adapter, TRADD, bridges the receptor with FADD. The signal transduction pathways for the two pro-apoptotic TRAIL receptors (DR4/TRAIL-R1 and DR5/TRAIL-R2/Apo-2/TRICK2/Killer) also seems to use both FADD and procaspase-8 in keeping with their molecular brethren¹⁹⁻²¹.

Death-inducing TNF-receptor family members and their cognate ligands serve many crucial physiological functions including tumour killing, lymphocyte culling and establishing zones of immune privilege. It is therefore not surprising that elaborate molecular control mechanisms have evolved to fine-tune this system to prevent either inappropriate or inadequate killing, and that disease pathologies can evolve when these systems fail. Among these controls, for example, are polypeptide modulators that serve largely to deactivate death-receptor activity, including decoy receptors that lack a carboxy-terminal 'death domain' and can function as ligand sinks to prevent engagement of the apoptotic pathway (for example, DcR1/TRAIL-R3/TRID and DcR2/TRAIL-R4/TRUNDD for TRAIL). Alternatively, recruitment of caspase-8 into the successfully ligated death-receptor complex is prevented by c-FLIP, a catalytically

incompetent pseudo-caspase that also desensitizes receptor-mediated apoptosis^{22,23}. Together, these and other regulatory elements (such as the extent of nuclear factor (NF)- κ B activation) determine the relative sensitivity or resistance of cells to ligand-provoked cell death, which clearly differ between normal and cancerous cells.

TRAIL therapeutic agent

The focus on TRAIL as a potential therapeutic agent became obvious as surprising differential sensitivity to TRAIL-stimulated apoptosis was observed between normal and cancerous cells^{24,25}. Approximately 80% of human cancer cell lines, representing colon, lung, breast, skin, kidney and brain tumours, are sensitive at least to some extent to TRAIL, whereas most normal cell types are relatively resistant. However, the molecular basis for this differential sensitivity is not clear. TRAIL and its receptors are broadly expressed in most organ systems, thus indicating the importance of the presence or absence of regulatory molecules in determining apoptotic sensitivity. Initially it seemed that the presence of TRAIL decoy receptors in normal cells (but not cancer cells) might confer resistance to TRAIL and thus explain this phenomenon²⁶, but a correlation of TRAIL sensitivity in cancer cells with reduced decoy-receptor mRNA expression cannot be firmly established²⁷. Other studies indicate that the subcellular distribution of the death and decoy receptors might account for the differential sensitivity of cells to TRAIL²⁸, or that c-FLIP levels dictate vulnerability to death receptor-mediated killing²⁹. Regardless of the apparent confusion in being able to explain the molecular basis for TRAIL sensitivity in cancer cells compared with resistance in normal cells, recombinant TRAIL has appealing therapeutic potential for the treatment of a variety of human cancers. An important consideration that distinguishes this approach from targeting Bcl-2 is that death receptor-mediated apoptosis in itself is predicted to be independent of p53 status, which is corrupted in about 50% of all primary human tumours³⁰. TRAIL-mediated apoptosis is also largely independent of Bcl-2 itself because it bypasses the 'intrinsic' cell death machinery in favour of a direct caspase activation pathway. In principle this should permit the treatment of chemoresistant cancers with TRAIL. Time will tell whether this is so. Recent evidence, for example, suggests that p53 is a transactivator of TRAIL receptor expression and thus might have a role in governing TRAIL sensitivity *in vivo*³¹.

Preclinical data

Nevertheless, promising results have been obtained in preclinical animal models involving human tumour xenografts into SCID or nude mice^{24,25,32}. These include mammary and colon carcinomas as well as intracranial gliomas. TRAIL was able to prevent the growth of evolving tumours immediately after xenotransplantation and, more importantly, decreased the size of established tumours in this model. For gliomas, injected recombinant TRAIL caused complete disease regression and entirely ablated tumour mass. For colon carcinomas the combination of subthreshold doses of recombinant TRAIL with existing chemotherapeutic agents resulted in a substantial positive interaction, completely eliminating the tumours in some animals. Taken together, these results look very promising, although in the absence of human clinical testing there is no certainty that the positive results achieved in animal models will translate to the complexities of the true human disease. One caveat, for example, is the potential adverse events that might accompany TRAIL injection into human patients. This has been largely addressed by preclinical safety studies in non-human primates (cynomolgus monkeys) that do not show adverse reactions to substantial doses of recombinant human TRAIL (10 mg kg⁻¹ d⁻¹ for 7 days)²⁴. One of the appealing features of TRAIL as a pro-apoptotic receptor ligand is that it does not seem to have the extreme liver toxicity that has precluded the testing *in vivo* of related death-inducing ligands such as CD95 ligand and TNF- α , which both cause massive haemorrhagic necrosis of various tissues including the liver. Although this has not been observed with TRAIL in diverse species from rodents to primates, a recent study indicates that human hepatocytes in culture might in fact be responsive to TRAIL and would thus predict TRAIL toxicity in humans^{33,34}. This is being cautiously evaluated by proponents of TRAIL therapy

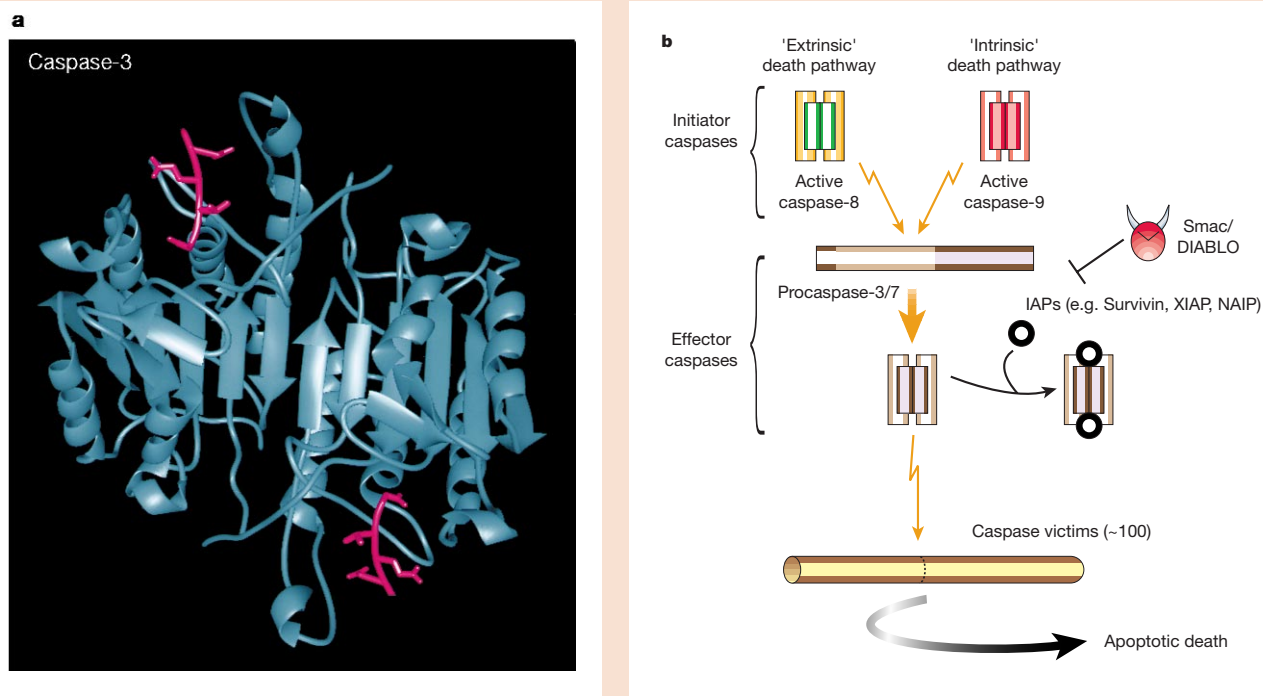


Figure 3 Caspases. Caspases are cysteinyl proteases that mediate most events that culminate in the apoptotic phenotype. Members of this protease family have common structural elements and are functionally redundant in some cases. **a**, The caspase-3 tetramer (blue) is composed of two large and two small subunits (derived from two proenzymes) generating two independent active sites that can be bound with inhibitors (magenta) to block catalytic activity. **b**, Apoptotic cell death is mediated by 'effector' caspases, such as caspase-3 and caspase-7, which cleave a limited subset of critical cellular polypeptides to manifest the apoptotic phenotype. These 'effector' caspases can be activated through proteolytic processing by upstream 'initiator' caspases such as caspase-8 and caspase-9. Two major activation pathways for these 'initiator' caspases are known, the 'intrinsic' pathway (see Fig. 1b) and the 'extrinsic' pathway (see Fig. 2b). Some degree of cross-talk between the two pathways seems to be mediated by tBid, a caspase-8-truncated form of the Bcl-2-related protein Bid (not shown). Active effector caspases are regulated by inhibitor-of-apoptosis (IAP) proteins, which block their catalytic activity and destine them for degradation. The mitochondrial cofactor protein Smac/DIABLO relieves this inhibition to facilitate full engagement of the proteolytic pathway.

(Genentech/Immunex) under conditions that would be representative of those planned for the clinic. It seems, for example, that not all recombinant TRAIL ligands are equal, as is predicted by the unexpected importance of Zn^{2+} ions in coordinating the trimeric organization of the functional ligand^{35,36}. Similarly, appropriate therapeutic windows need to be established, and all potential adverse events predicted by preclinical studies will require careful monitoring once TRAIL therapy is brought to the clinic. However, at the same time the balance of pros and cons needs to be considered because the benefits for cancer patients could potentially outweigh the disadvantages by a wide margin.

Caspases as therapeutic targets

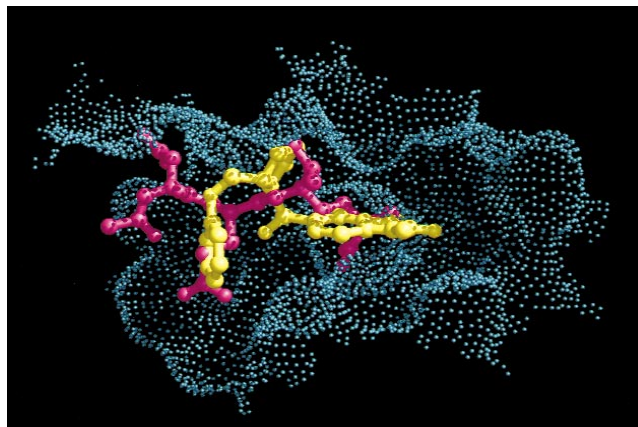
Every mammalian cell contains a complement of different caspase proteases that suitably equip it for apoptotic death (Fig. 3a). Although there is emerging evidence that sustained pro-apoptotic stress results in the transcriptional upregulation of these proteases, most cell deaths are dependent on the pre-existing caspase armament within each cell. In this way, cells can engage the cell death pathway at will without requiring elaborate biochemical procedures (for example, transcription and translation) that might be subject to interference (for example, by metabolic deficiency or viral pathogens). That cells do not spontaneously commit suicide is because these proteases are dormant when in their unprocessed proenzyme state. Caspase catalytic activity is associated with proteolytic maturation, which is launched through either the 'extrinsic' or 'intrinsic' pathways (see Figs 1b, 2b). However, as might be expected, this poised apoptotic system sometimes misfires, leading to premature cell death. This seems to happen often after acute cellular injury but is also prevalent in chronic disorders where the pro-apoptotic stimulus is obscure. In these cases, caspase inhibition has shown extraordinary promise in multiple disease models, particularly for acute indications.

Caspases were probably one of the first obvious therapeutic targets for modulating apoptosis and they remain the most viable approach to blocking apoptotic cell death as opposed to the death-inducing capacities of the preceding examples, Bcl-2 antisense and TRAIL. In fact, caspase inhibitor programmes were well underway in pharmaceutical companies before the discovery that these proteases had a central role in apoptosis. Their target was the interleukin-1 β -converting enzyme (ICE, now designated caspase-1 (ref. 37)), a novel cysteine protease that is responsible for the proteolytic maturation and concomitant activation of pro-inflammatory cytokines, including pro-interleukin-1 β and IGIF/interleukin-18 (refs 38, 39). Soon after the discovery that one of the gene products that was essential for apoptosis in *C. elegans* was an ICE-related protease (CED-3)⁴⁰, the role of caspases in mammalian apoptosis was rapidly unveiled with the use of the knowledge accumulated with ICE as a template. For example, inhibitors that were designed to target one of these proteases (caspase-3) prevented apoptosis as gene disruption did for neurons in caspase-3-null mice^{41,42}. So far, 14 mammalian caspases have been identified, of which 12 human orthologues are known^{43,44}. Functionally, these proteases divide into two major subfamilies: those related to ICE (caspase-1, caspase-4 and caspase-5) function predominantly, if not exclusively, in cytokine maturation; the remainder perform an elaborated CED-3 function to mediate apoptosis. Among these latter CED-3-like mammalian caspases, a further subdivision exists: 'initiator' caspases (for example, caspase-8, caspase-9 and caspase-10) respond to pro-apoptotic stimuli and subsequently catalyse the activation of more abundant, and catalytically robust 'effector' caspases (for example caspase-3 and caspase-7) that largely perform the proteolytic cleavage events necessary to mediate the apoptotic phenotype (Fig. 3b). This division of duties is consistent with the proteolytic specificity of individual

Box 1

Caspase inhibitor design

Caspase inhibitor design has taken advantage of key features found within the active-site substrate-binding cleft as revealed by X-ray crystallography (the active site 'bite' of caspase-3 is shown in the figure opposite with overlaid peptoid (magenta) and isatin (yellow) inhibitors). It was recognized early on that extremely potent inhibitors for these enzymes could be made by coupling an appropriate P₁-Asp tetrapeptide (tetrapeptides are sufficient for specific recognition by these proteases and Asp in the P₁ position is a near absolute requirement) with an appropriate C-terminal electrophile, such as an aldehyde or ketone, to attack the thiol side chain of the active-site cysteine residue (at the right side of the magenta inhibitor). From this, inhibitors have evolved along at least two routes to circumvent the disastrously poor cell permeability that preclude these molecules from utility except *in vitro*. Peptoid inhibitors retain the physical characteristics of the original substrate counterpart, including high potency and specificity, but can bypass many of the liabilities of their peptide predecessors. Non-peptide inhibitors, such as isatins⁶³, do not contain peptide equivalents (not even the Asp carboxylate or a P₁ equivalent) but their highly electrophilic nature and metabolic instability place major limitations on their current utility. Overall, major gains have been made in the cell permeability of these next-generation peptoid and isatin inhibitors, making them 100–1000-fold more potent than z-VAD-like inhibitors in cell-based apoptosis assays and yielding substantial improvements *in vivo* as a consequence. So far, isatins are effective only against the effector caspases (caspase-3 and caspase-7).



caspase family members⁴⁵ and indicates that the functional redundancies built into each of the three caspase subgroups, coupled with the cell-type specificity that has evolved in complex multicellular organisms, might permit the specific targeting of individual caspases for therapeutic purposes.

Caspase inhibitors

Although a caspase inhibitor has yet to reach the clinic for the inhibition of apoptosis (ICE inhibitors have done so recently for the treatment of rheumatoid arthritis), preclinical studies are compelling, and the plethora of diseases in which they might have efficacy makes their therapeutic potential enormous. Therapeutics are being developed for various caspase family members, with the most attention being paid to caspase-3 as a major contributor to the apoptotic machinery in many cell types.

With few exceptions, all of the proof-of-concept preclinical studies with caspase inhibitors in animal models of human diseases have been performed with active-site mimetic peptide ketones (for example, benzyloxycarbonyl (z)-VAD-fluoromethylketone (fmk), z-YVAD-fmk/chloromethylketone (cmk), z-DEVD-fmk/cmk and z-D-cmk)⁴⁶. These molecules are all relatively non-selective caspase inhibitors and although they are not appropriate tools for dissecting out the contributions made by individual caspase family members to the apoptotic response that occurs in disease models, they have provided extremely valuable preclinical insight into the potential that caspase inhibition might eventually have in humans. For example, in at least five different models of ischaemia–reperfusion injury (liver, cardiac, renal, intestinal and cerebral), caspase inhibition has shown remarkable efficacy^{47–51}. In addition to decreased apoptosis, caspase inhibition improved survival (for example threefold in liver ischaemia), decreased infarct volumes (by 50% in both cardiac and focal cerebral ischaemia) and, ultimately, markedly improved organ function (for example renal function and neurodeficit in models of kidney and focal cerebral ischaemia, respectively). These findings partly address the critical question of whether cells that are saved by anti-apoptotic therapies retain function. Importantly, many of these inhibitor studies (for example, cardiac, renal and cerebral ischaemia) demonstrated that caspase inhibitors that were administered after the ischaemic insult, and coincident with reperfusion, retained their efficacy owing to the natural delay in the apoptotic response, an important practical issue that distinguishes this therapeutic strategy from many that have preceded it. Taken together, these studies

demonstrate that caspase inhibition should have protective value in organ transplantation, cardiac arrest and stroke.

In addition, caspase inhibition has shown promise in preclinical animal models of other disorders of the by decreasing apoptotic cell death, including traumatic brain injury, status epilepticus, amyotrophic lateral sclerosis (ALS) and Parkinson's disease^{52–55}. In traumatic injury, improved neurological recovery accompanied apoptosis inhibition and in a mutant-superoxide-dismutase-1 model for ALS there was a significant delay in disease onset and mortality. Caspase inhibitors have also shown efficacy in animal models of infectious diseases, including bacterial meningitis and sepsis^{56–58}. In the former, inhibitors blocked the apoptotic loss of hippocampal neurons, decreased spongiform encephalopathy and markedly improved behavioural scores. In the latter, septic animals more than doubled their survival probability. It therefore seems likely that the acute use of caspase inhibitors will have measurable clinical benefits in addition to a general inhibition of apoptosis. For chronic use, the issues will be much more complex. Although somewhat controversial, it seems that the neuronal loss in chronic neurodegenerative disorders is apoptotic; moreover, caspases might themselves have a role in exacerbating disease pathogenesis, as has been suggested for Huntington's and Alzheimer's diseases^{59–61}. Thus, caspase inhibition shows immense preclinical promise, which now requires usable therapeutics that can be taken forward into human clinical trials.

Potent specific caspase inhibitors that are able to cross the membrane of cells are being developed by various pharmaceutical companies (Merck, SmithKline Beecham, BASF, Idun/Novartis and Vertex) (see Box 1), although very little is known about their preclinical status or clinical agendas. Regardless of this apparent information blackout, the remarkable efficacy that 'old generation' caspase inhibitors have in animal models indicates that the emerging crop should indeed be impressive and will no doubt be viable in the clinic. It is also clear that the first entries will probably target acute apoptotic injuries such as cerebral stroke, trauma-induced neurodegeneration, cardiac ischaemia–reperfusion injury, transplantation, acute liver injury and sepsis. Finally, a novel gene-therapy approach has been suggested for the treatment of HIV-mediated AIDS⁶². In this approach, the maturation sites within the caspase-3 polypeptide are replaced with HIV protease recognition motifs. Cells that are infected by the virus undergo apoptosis when the protease clips the engineered caspase, thereby selectively deleting infected cells.

Whether this strategy will add value to the current armament of AIDS therapies is unknown, but it is representative of the many prospects for caspase-modulating therapeutics.

Future prospects

Is the ability to modulate the life or death of a cell — whether it be by using small organic molecules, biological agents, antisense oligonucleotides or gene therapy — indeed ‘too good to be true’? Although these are still early days, it is difficult not to get excited about the significant advances that have already been made. The true therapeutic benefit of apoptosis modulation for the treatment of some of the most devastating human diseases remains to be discovered. As the biochemical and molecular complexities of the apoptotic pathway are elucidated, new therapeutic strategies will no doubt arise and more obvious paths forwards for the modulation of chronic apoptosis will appear. Apoptosis-based therapeutics are clearly within our grasp, but the emerging crop are likely to be just the tip of the iceberg. □

- Ellis, R. E., Yuan, J. Y. & Horvitz, H. R. Mechanisms and functions of cell death. *Annu. Rev. Cell Biol.* **7**, 663–698 (1991).
- Wang, J. L. *et al.* Structure-based discovery of an organic compound that binds bcl-2 protein and induces apoptosis of tumor cells. *Proc. Natl Acad. Sci. USA* **97**, 7124–7129 (2000).
- Adams, J. M. & Cory, S. The Bcl-2 protein family: arbiters of cell survival. *Science* **281**, 1322–1326 (1998).
- Reed, J. C. Bcl-2 family proteins. *Oncogene* **17**, 3225–3236 (1998).
- Veis, D. J., Sorenson, C. M., Shutter, J. R. & Korsmeyer, S. J. Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair. *Cell* **75**, 229–240 (1993).
- Cory, S. Regulation of lymphocyte survival by the bcl-2 gene family. *Annu. Rev. Immunol.* **13**, 513–543 (1995).
- Schlagbauer-Wadl, H. *et al.* Bcl-2 antisense oligonucleotides (G3139) inhibit Merkel cell carcinoma growth in SCID mice. *J. Invest. Dermatol.* **114**, 725–730 (2000).
- Jansen, B. *et al.* bcl-2 antisense therapy chemosensitizes human melanoma in SCID mice. *Nature Med.* **4**, 232–234 (1998).
- Jansen, B. *et al.* Bcl-2 antisense plus dacarbazine therapy for malignant melanoma. *Proc. Am. Assoc. Cancer Res. Conf. Programmed Cell Death Regul.* A59 (2000).
- Waters, J. S. *et al.* Phase I clinical and pharmacokinetic study of bcl-2 antisense oligonucleotide therapy in patients with non-Hodgkin's lymphoma. *J. Clin. Oncol.* **18**, 1812–1823 (2000).
- Zangemeister-Wittke, U. *et al.* A novel bispecific antisense oligonucleotide inhibiting both bcl-2 and bcl-xL expression efficiently induces apoptosis in tumor cells. *Clin. Cancer Res.* **6**, 2547–2555 (2000).
- Reed, J. C. Splicing and dicing apoptosis genes. *Nature Biotechnol.* **17**, 1064–1065 (1999).
- Taylor, J. K., Zhang, Q. Q., Wyatt, J. R. & Dean, N. M. Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. *Nature Biotechnol.* **17**, 1097–1100 (1999).
- Que, F. G. *et al.* Cholangiocarcinomas express Fas ligand and disable the Fas receptor. *Hepatology* **30**, 1398–1404 (1999).
- Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* **281**, 1305–1308 (1998).
- Schneider, P. & Tschopp, J. Apoptosis induced by death receptors. *Pharm. Acta Helv.* **74**, 281–286 (2000).
- Boldin, M. P., Goncharov, T. M., Goltsev, Y. V. & Wallach, D. Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* **85**, 803–815 (1996).
- Muzio, M. *et al.* FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
- Kischkel, F. C. *et al.* Apo2L/TRAIL-dependent recruitment of endogenous FADD and caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611–620 (2000).
- Sprick, M. R. *et al.* FADD/MORT1 and caspase-8 are recruited to TRAIL receptors 1 and 2 and are essential for apoptosis mediated by TRAIL receptor 2. *Immunity* **12**, 599–609 (2000).
- Bodmer, J. L. *et al.* TRAIL receptor-2 signals apoptosis through FADD and caspase-8. *Nature Cell Biol.* **2**, 241–243 (2000).
- Thome, M. *et al.* Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. *Nature* **386**, 517–521 (1997).
- Rasper, D. M. *et al.* Cell death attenuation by ‘Usurpin’, a mammalian DED-caspase homologue that precludes caspase-8 recruitment and activation by the CD-95 (Fas, APO-1) receptor complex. *Cell Death Differ.* **5**, 271–288 (1998).
- Ashkenazi, A. *et al.* Safety and antitumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155–162 (1999).
- Walczak, H. *et al.* Tumor necrosis factor-related apoptosis-inducing ligand in vivo. *Nature Med.* **5**, 157–163 (1999).
- Marsters, S. A., Pitti, R. A., Sheridan, J. P. & Ashkenazi, A. Control of apoptosis signaling by Apo2 ligand. *Recent Prog. Horm. Res.* **54**, 225–234 (1999).
- Griffith, T. S. & Lynch, D. H. TRAIL: a molecule with multiple receptors and control mechanisms. *Curr. Opin. Immunol.* **10**, 559–563 (1998).
- Zhang, X. D., Franco, A. V., Nguyen, T., Gray, C. P. & Hersey, P. Differential localization and regulation of death and decoy receptors for TNF-related apoptosis-inducing ligand (TRAIL) in human melanoma cells. *J. Immunol.* **164**, 3961–3970 (2000).
- Kim, K., Fisher, M. J., Xu, S. Q. & el-Deiry, W. S. Molecular determinants of response to TRAIL in killing of normal and cancer cells. *Clin. Cancer Res.* **6**, 335–346 (2000).
- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49–53 (1991).
- Takimoto, R. & el-Deiry, W. S. Wild-type p53 transactivates the KILLER/DR5 gene through an intronic sequence-specific DNA-binding site. *Oncogene* **19**, 1735–1743 (2000).
- Ashkenazi, A. The Apo2L/TRAIL system: therapeutic opportunities. *Proc. Am. Assoc. Cancer Res. Conf. Programmed Cell Death Regul.* (2000).
- Nagata, S. Steering anti-cancer drugs away from the TRAIL. *Nature Med.* **6**, 502–503 (2000).
- Jo, M. *et al.* Apoptosis induced in normal human hepatocytes by tumor necrosis factor-related apoptosis-inducing ligand. *Nature Med.* **6**, 564–567 (2000).
- Hymowitz, S. G. *et al.* A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633–640 (2000).
- Hymowitz, S. G. *et al.* Triggering cell death: the crystal structure of Apo2L/TRAIL in a complex with death receptor 5. *Mol. Cell* **4**, 563–571 (1999).
- Alnemri, E. S. *et al.* Human ICE/CED-3 protease nomenclature. *Cell* **87**, 171 (1996).
- Thornberry, N. A. *et al.* A novel heterodimeric cysteine protease is required for interleukin-1 β processing in monocytes. *Nature* **356**, 768–774 (1992).
- Cerretti, D. P. *et al.* Molecular cloning of the interleukin-1 β converting enzyme. *Science* **256**, 97–100 (1992).
- Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. & Horvitz, H. R. The *C. elegans* cell death gene *ced-3* encodes a protein similar to mammalian interleukin-1 β -converting enzyme. *Cell* **75**, 641–652 (1993).
- Nicholson, D. W. *et al.* Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis. *Nature* **376**, 37–43 (1995).
- Kuida, K. *et al.* Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* **384**, 368–372 (1996).
- Nicholson, D. W. Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death Differ.* **6**, 1028–1042 (1999).
- Earnshaw, W. C., Martins, L. M. & Kaufmann, S. H. Mammalian caspases: structure, activation, substrates, and functions during apoptosis. *Annu. Rev. Biochem.* **68**, 383–424 (1999).
- Thornberry, N. A. *et al.* A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J. Biol. Chem.* **272**, 17907–17911 (1997).
- Garcia-Calvo, M. *et al.* Inhibition of human caspases by peptide-based and macromolecular inhibitors. *J. Biol. Chem.* **273**, 32608–32613 (1998).
- Cursio, R. *et al.* A caspase inhibitor fully protects rats against lethal normothermic liver ischemia by inhibition of liver apoptosis. *FASEB J.* **13**, 253–261 (1999).
- Mocanu, M. M., Baxter, G. F. & Yellon, D. M. Caspase inhibition and limitation of myocardial infarct size: protection against lethal reperfusion injury. *Br. J. Pharmacol.* **130**, 197–200 (2000).
- Farber, A. *et al.* A specific inhibitor of apoptosis decreases tissue injury after intestinal ischemia–reperfusion in mice. *J. Vasc. Surg.* **30**, 752–760 (1999).
- Daemen, M. A. *et al.* Inhibition of apoptosis induced by ischemia–reperfusion prevents inflammation. *J. Clin. Invest.* **104**, 541–549 (1999).
- Endres, M. *et al.* Attenuation of delayed neuronal death after mild focal ischemia in mice by inhibition of the caspase family. *J. Cereb. Blood Flow Metab.* **18**, 238–247 (1998).
- Yakovlev, A. G. *et al.* Activation of CPP32-like caspases contributes to neuronal apoptosis and neurological dysfunction after traumatic brain injury. *J. Neurosci.* **17**, 7415–7424 (1997).
- Kondratyev, A. & Gale, K. Intracerebral injection of caspase-3 inhibitor prevents neuronal apoptosis after kainic acid-evoked status epilepticus. *Brain Res. Mol. Brain Res.* **75**, 216–224 (2000).
- Li, M. *et al.* Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model. *Science* **288**, 335–339 (2000).
- Schierle, G. S. *et al.* Caspase inhibition reduces apoptosis and increases survival of nigral transplants. *Nature Med.* **5**, 97–100 (1999).
- Braun, J. S. *et al.* Neuroprotection by a caspase inhibitor in acute bacterial meningitis. *Nature Med.* **5**, 298–302 (1999).
- Grobmyer, S. R. *et al.* Peptidomimetic fluoromethylketone rescues mice from lethal endotoxin shock. *Mol. Med.* **5**, 585–594 (1999).
- Hotchkiss, R. S. *et al.* Prevention of lymphocyte cell death in sepsis improves survival in mice. *Proc. Natl Acad. Sci. USA* **96**, 14541–14546 (1999).
- Sanchez, I. *et al.* Caspase-8 is required for cell death induced by expanded polyglutamine repeats. *Neuron* **22**, 623–633 (1999).
- Goldberg, Y. P. *et al.* Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nature Genet.* **13**, 442–449 (1996).
- Gervais, F. G. *et al.* Involvement of caspases in proteolytic cleavage of Alzheimer's amyloid-beta precursor protein and amyloidogenic A β peptide formation. *Cell* **97**, 395–406 (1999).
- Vocero-Akbani, A. M., Heyden, N. V., Lissy, N. A., Ratner, L. & Dowdy, S. F. Killing HIV-infected cells by transduction with an HIV protease-activated caspase-3 protein. *Nature Med.* **5**, 29–33 (1999).
- Lee, D. *et al.* Potent and selective nonpeptide inhibitors of caspases 3 and 7 inhibit apoptosis and maintain cell functionality. *J. Biol. Chem.* **275**, 16007–16014 (2000).

Acknowledgements

I thank S. Roy, S. Xanthoudakis, D. McKay, C. Bayly and K. Clark (Merck) and A. Ashkenazi (Genentech) for assistance with this manuscript and its figures.

The main attraction

Separating the blood, stirring the pot, using magnets.

Spindrive

Bel-Art Products www.bel-art.com

Turn a magnetic stirrer into an orbital shaker

Place the Scienceware Spindrive Orbital Shaker Platform onto any standard magnetic stirrer and the rotation of the stirrer will activate the orbital motion of the shaker platform. Speed adjustment is accomplished using the speed setting on the magnetic stirrer. The Spindrive Platform uses no electricity or batteries, only magnetic power. The Spindrive Orbital Shaker Platform is 225 mm square with a rotating orbit diameter of 21.7 mm. The economical, compact design saves bench-top workspace. The platform also features a non-skid surface to prevent samples from falling off.

Reader Service No. 100

Stir Light

Barnstead/Thermolyne
www.barnsteadthermolyne.com

Stirring news

The Stir Light illuminated top stirrer has been redesigned with a new cast-aluminium housing for greater durability and drip-edge channels to keep spills away from the controls. An integral support rod holder has been added for increased functionality. The Thermolyne Stir Light provides magnetic coupling suitable for stirring viscous aqueous solutions, and the illuminated top facilitates the observation of titration endpoints. Stirring speed range is 100–1,000 r.p.m.

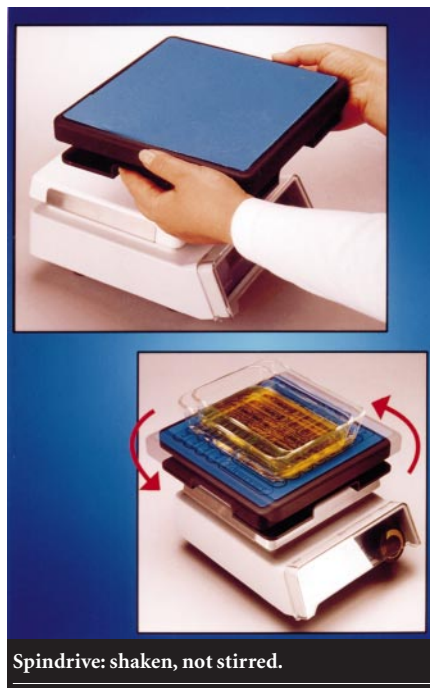
Reader Service No. 101

MACS LD Column

Miltenyi Biotec
www.miltenyibiotec.com

Magnetic cell separation of virtually any cell type

The LD Column completes the range of separation columns from Miltenyi Biotec for magnetic cell separation. Each column type is optimized for a certain range of total cells and magnetically labelled cells. MACS LD columns are optimized for convenient depletion of virtually any cell type. Starting from up to 5×10^8 total cells, a maximum of 1×10^8 cells can be depleted. Cells are magnetically labelled with MACS MicroBeads and the cell suspension is passed over the LD Col-



Spindrive: shaken, not stirred.

umn and washed. Highly pure cell populations are collected as effluent. The manufacturers claim that a depletion efficiency of up to 3 log can be achieved. The flow stop LD Columns, which don't require any additional equipment such as 3-way stop cocks or flow resistors, show a consistent flow rate and are compatible with the MACS separators Midi-MACS, Vario-MACS and SuperMACS.

Reader Service No. 102

MagnaBot

Promega www.promega.com

High-volume separations

The MagnaBot Magnetic Separation Device is designed for high-throughput bioseparation using magnetic particles such as Promega's MagneSil paramagnetic particles. MagnaBot is compatible with various instrument platforms including the Beckman BioMek 2000 Laboratory Automation Workstation, and with 96-well plates including Dynatech 001-010-2701 V-bottom and Costar 43119 Deep Well. It has been successfully used in protocols for plasmid and PCR product purification. The device is built for ease of use: a plastic 96-well microtitre plate is placed over the 96-well magnet such that each of the 24 magnetic pins resides in the interstices between four wells. The neodymium iron boron magnets attract the magnetic silica or particles to the side and away from the bottom of each well in seconds, allowing the removal of particle-free fluid.

Reader Service No. 103

MPG

CPG, Inc. www.cpg-biotech.com

Magnetic secondary antibody particles

MPG secondary antibody particles are Magnetic Porous Glass particles to which affinity-purified secondary antibodies have been covalently attached. The nominal binding capacity of 1 mg of these particles is ≥ 150 pg of immunoglobulin. MPG particles bind immunoglobulins from various species, including mouse, human, rabbit, rat, goat and sheep primary antibodies. MPG secondary antibody particles eliminate centrifugation for easy scale-up and automation, and give flexibility in assay configuration. MPG secondary antibody particles can be used for immunoassays, separation of unbound labelled tracers, small-scale immunoaffinity purification and immunoprecipitation reactions.

Reader Service No. 104

Human Total RNA

Stratagene www.stratagene.com

RNA for a wide range of uses

Stratagene now offers a comprehensive list of total RNA from normal and diseased tissues for various lines of investigation into gene discovery and functionality. Well-documented donor and tissue pathology information is provided with all of these RNA products. Matched single donor sets of total RNA from normal and diseased tissue are now available for microarray interrogation studies. Human total RNA is supplied in 100- μ g aliquots conveniently stored in 0.1 mM EDTA, pH 8.0.

Reader Service No. 105

GX-1

TEAC America, Inc. www.teac.com

Portable data acquisition and recording

GX-1 is a lightweight, portable, data-acquisition and -recording system combining the data-storage capacity and wide frequency response of a data recorder with the programmable sampling rates and integrated signal-conditioning amplifiers of a data-acquisition system in a very compact package. GX-1 is a full-featured 16-channel instrument offering continuous recording at bandwidths up to 200 kHz per channel. It is suitable for any mobile, multi-channel application such as automotive, off-road vehicles, military and other test applications where portability and absolute data integrity and accuracy are required. Using the internal 256-MB DRAM as a buffer, it can be used as a front-end for a PC, downloading via a fast SCSI-2 interface. As a stand-alone recorder, the GX-1 provides flexible storage



Attractive proposition: the LCD column.

new on the market



GX-1: get it taped.

options of an internal 25-GB AIT tape drive, magneto-optical drive, or PCMCIA memory card.

Reader Service No. 106

Slide-A-Lyzer Mini Unit

Pierce www.piercenet.com

Dialyse microlitre samples in minutes

Slide-A-Lyzer Mini Dialysis Units are designed for sample volumes as small as 10 µl. Sample recovery is claimed to be high, at 9–10 µl for a 10-µl sample. A 100-µl, pH 2.8 buffer sample can be converted to pH 9.4 dialysing against a litre of bicarbonate buffer, pH 9.4, in less than 10 minutes — some 'home-made' devices might take several hours. The Mini Units are made with low protein and nucleic-acid-absorbing regenerated cellulose membranes in 3.5K, 7K or 10K molecular-weight cutoffs in packs of 50 or economically 250 units.

Reader Service No. 107

Bio-Ferrograph 2100

Guilfoyle, Inc.

Separating whole blood cells magnetically

Produced by Guilfoyle, Inc., of Belmont, Massachusetts (tel. 617-489-5007), the Bio-Ferrograph 2100 uses cell-capture cytometry to capture cells with one-cell-in-a-million

selectivity and sensitivity, and since only disposable cassettes and syringes contact the samples it eliminates cross-contamination. The user simply incubates each sample with an appropriate monoclonal antibody conjugated to submicrometre magnetic particles, then adds the samples to reservoirs on the instrument and selects a program. Fitted with five 1-ml sample reservoirs, the Bio-Ferrograph 2100 can process a 0.3-ml sample in each reservoir within 20 minutes. The selected cells are available on a glass slide for observation and the cells from five samples can be deposited simultaneously within bracketed areas on a single slide. Providing a short processing flow path, cells can be captured from whole blood samples without removing the red blood cells.

Reader Service No. 108

inoLab series

WTW www.wtw-inc.com

Three-stage approach to measurement

The inoLab series of instruments from WTW Measurement Systems, Inc. come in three levels of configuration. Models are available to measure pH/mV, ORP, ISE, temperature, dissolved oxygen conductivity or salinity. inoLab Level 1 meters are configured for ease of use and reliability on a budget. Standard features include automatic calibration, temperature compensation and measurement storage. The instrument is battery powered with an optional mains adapter. Additional features on Level 2 include: bidirectional RS232 interface, integrated datalogger and optional built-in printer. The top-of-the-line Level 3 range includes a flexible and powerful line of pH/mV, oxygen, conductivity or multiparameter (pH/mV, conductivity, dissolved oxygen) meters. Level 3 series meters also offer additional accuracy and unlimited upgradeability.

Reader Service No. 109



Bio-Ferrograph 2100: one in a million.

CollagenCult

StemCell Technologies

www.stemcell.com

Colonial media

CollagenCult is a line of collagen-based media optimized for the growth of human or murine haematopoietic colonies (BFU-E, CFU-GM, and CFU-GEMM). Unlike methylcellulose-based media, CollagenCult allows the colony assays to be fixed and stained with antibodies against specific lineage markers, enabling the definitive identification of colony types that may otherwise be difficult to distinguish from each other. The fixed and stained slides can be kept indefinitely to provide a permanent record of the assays for future reference.

Reader Service No. 110

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

TREVIGEN

new

PARP

**Poly (ADP-ribose)
Polymerase Reagents**

- PARP inhibition assay
- Biotinylated NAD
- Recombinant human PARP
- Anti PARP and PAR

www.trevigen.com

US and Canada: 1 800 873 8443
TEL: 301 216 2800 • FAX: 301 216 2801

READER ENQUIRY NO. 50

Custom Peptides & Polyclonal Antibodies

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics
U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 53

Look no further

for the best value in

Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA (+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
GLX055 email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33